

NPS55-78-034

NAVAL POSTGRADUATE SCHOOL

Monterey, California



AN INTERACTIVE SOFTWARE PACKAGE
FOR TIME SERIES ANALYSIS

by

F. Russell Richards

and

Stephen R. Woodall

November 1978

Approved for public release; distribution unlimited.

FEDDOCS
D 208.14/2:
NPS-55-78-034

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral T. F. Dedman
Superintendent

J. R. Borsting
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-78-034	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Interactive Software Package for Time Series Analysis		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) F. Russell Richards and Stephen R. Woodall		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		12. REPORT DATE November 1978
		13. NUMBER OF PAGES 171
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Time series FORTRAN Time Series Editor Box-Jenkins		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An expanded package of interactive FORTRAN computer programs has been developed for the analysis and forecasting of time series data. The package, called the Time Series Editor, is designed to employ the iterative Box-Jenkins methodology of time series analysis. The Time Series Editor was developed for time-shared use on the Control Program/Cambridge Monitor system (CP/CMS) at the U. S. Naval Postgraduate School, but can be modified for use on other time-sharing systems with a FORTRAN capability. The		

20. ABSTRACT Continued.

Time Series Editor assists in data preparation and entry, analysis, modeling, forecasting and diagnostic testing. Utilization of the package, following the included User's Guide, requires only a limited knowledge of the computer system, with all required user responses interactively prompted by the Editor.

ABSTRACT

An expanded package of interactive FORTRAN computer programs has been developed for the analysis and forecasting of time series data. The package, called the Time Series Editor, is designed to employ the iterative Box-Jenkins methodology of time series analysis. The Time Series Editor was developed for time-shared use on the Control Program/Cambridge Monitor System (CP/CMS) at the U.S. Naval Postgraduate School, but can be modified for use on other time-sharing systems with a FORTRAN capability. The Time Series Editor assists in data preparation and entry, analysis, modeling, forecasting and diagnostic testing. Utilization of the package, following the included User's Guide, requires only a limited knowledge of the computer system, with all required user responses interactively prompted by the Editor.

TABLE OF CONTENTS

I.	INTRODUCTION -----	9
II.	BOX-JENKINS METHODOLOGY: A PRIMER -----	16
	A. LISTING OF NOTATION -----	16
	B. TIME SERIES, STOCHASTIC MODELS AND STATIONARITY -----	17
	C. THE BASICS OF DIFFERENCE OPERATORS AND LINEAR FILTER MODELS -----	19
	D. HOMOGENEOUS NON-STATIONARY AND SEASONAL PROCESSES -----	22
	E. GENERAL CLASSES OF BOX-JENKINS MODELS -----	24
	F. THE AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTIONS -----	30
	G. COMMENTS ON MODEL SELECTION EMPLOYING AUTOCORRELATIONS AND PARTIAL AUTOCORRE- LATIONS -----	35
	H. ESTIMATION OF MODEL PARAMETERS -----	42
	I. DIAGNOSTIC CHECKING OF MODELS -----	44
	J. FORECASTING -----	46
III.	GENERALIZED ARIMA MODEL PARAMETRIC ESTIMATION --	51
	A. MATHEMATICAL DISCUSSION OF THE MARQUARDT ALGORITHM -----	52
	B. MODEL INPUT INFORMATION -----	53
	C. ALGORITHM DESCRIPTION -----	54
	D. MODEL OUTPUT AND DIAGNOSTIC CHECKS -----	60
IV.	DESCRIPTION OF THE TIME SERIES EDITOR -----	63
	A. THE EXECUTIVE PROGRAM -----	63
	B. DATA AND FILE MANAGEMENT -----	65
	1. OFFLINE Data Entry -----	65
	2. ZFORMAT Program -----	66

3.	Interactive Keyboard Input -----	70
4.	Output -----	71
5.	CMSWORK Program -----	72
C.	TRANS PROGRAM -----	72
D.	DIFF PROGRAM -----	74
E.	PLOT PROGRAM -----	75
F.	AUTO PROGRAM -----	75
G.	ESTIMATE PROGRAM -----	76
H.	YESTSEAS PROGRAM -----	77
I.	WMARQRTD PROGRAM -----	78
J.	XSUMSQ PROGRAM -----	79
K.	FORECAST PROGRAM -----	79
L.	ROOTS PROGRAM -----	80
M.	SIMULATE AND GENERATE PROGRAMS -----	81
N.	HELP OPTION -----	82
O.	SUMMARY -----	83
V.	EXAMPLE NON-SEASONAL TIME SERIES ANALYSIS -----	84
VI.	EXAMPLE SEASONAL TIME SERIES ANALYSIS -----	88
VII.	SUMMARY AND RECOMMENDATIONS -----	93
APPENDIX A:	User's Guide to the Time Series Editor ---	95
APPENDIX B:	Keyboard Session and Output from Analysis of Time Series G (Seasonal) -----	119
APPENDIX C:	Listing of Data Sets Residing in the Time Series Editor -----	160
LIST OF REFERENCES	-----	165
INITIAL DISTRIBUTION LIST	-----	167

LIST OF FIGURES

Figure 1.	The Box-Jenkins Forecasting Methodology -----	12
Figure 2.	Representation of a Time Series as the Output from a Linear Filter -----	21
Figure 3.	Example of Auto and Pauto Plots -----	36
Figure 4.	Auto and Pauto vs. Log for Mixed ARMA(p,q), with Seasonality Remaining of Period $s = 6$ -----	42
Figure 5.	Card Deck Arrangement for OFFLINE READ of Time Series Input Data in FORMAT(5F15.6) -----	67
Figure 6.	Card Deck Arrangement for OFFLINE READ of Time Series Input Data in Any Format -----	68
Figure 7.	Sample User Session with ZFORMAT Program ----	69
Figure 8.	Sample User Session with CMSWORK Program ----	73

LIST OF TABLES

Table I.	Basic Difference Operators	-----	20
Table II.	Properties of Basic ARMA Processes	-----	41

I. INTRODUCTION

Operations researchers, military planners and programmers, statisticians, economists, marketing personnel, managers and others are often faced with the need to analyze data in the form of a time series, which can be thought of as a sequence of observations of either a deterministic or stochastic process. In most cases, the objective of the analysis is to determine patterns or other recognizable behavior apparent in the data, and to then formulate a suitable mathematical model for the time series from which forecasts of future behavior can be obtained. The value to a decision maker of being able to predict the future with some reasonable and statistically quantifiable degree of accuracy can not be overstated. For example, in areas such as budget expenditures, recruiting performance, commodity prices, population levels, resource consumption, manpower levels and consumer demand for products, decision makers charged with planning for the future should base their decisions at least in part on the best available predictions about the future behavior of the time series in question.

Until the late 1960's, the techniques employed in time series analysis were primarily those of spectral analysis, with applications of harmonic analysis and mathematical transform theory. For an introduction to these topics, the reader is referred to Anderson [Ref. 2], Lewis [Ref. 7] and Jenkins [Ref. 10]. Due to the high degree of mathematical

sophistication required by the spectral analysis approach, the capability to perform time series analysis resided nearly exclusively with mathematicians and electrical engineers. As a consequence, the majority of decision makers came to use more understandable (and far less powerful) methods such as simple moving averages or exponential smoothing. However, in the late 1960's and early 1970's, the statistical analysis of time series by the methods developed by Box and Jenkins [Ref. 4] has gained widespread acceptance. Although these methods are mathematically nearly equivalent to the spectral approach (there are transformations that interconnect the two methodologies), the Box-Jenkins approach is described in vocabulary more familiar to operations researchers, statisticians, economists and managers, and is therefore being used more and more by them in building models from measurements of the past.

Many algorithms and computer programs for performing the analyses required by the Box-Jenkins approach have been developed and are available from several sources. Perhaps the best source is the collection of FORTRAN computer subroutines which resides in the International Mathematical and Statistical Library (IMSL) [Ref. 9]. The primary problem with using the available computer resources lies not in any deficiency of the programs themselves, but with the very nature of the Box-Jenkins methodology itself. The Box-Jenkins method is an iterative approach to time series

analysis, which is briefly described in Figure 1. [See Wheelwright and Makridakis, Ref. 16.]

As indicated by Figure 1, the Box-Jenkins method is a multi-stage, iterative process. It begins with the postulation of a general class of time series models which has been found, experimentally, to be extremely rich. Thereafter, the procedure continues as a trial-and-error process, with decision points where the analyst is required to select the next direction based on the best information available to him. Since each stage of the process outlined in Figure 1 may consist of several sub-steps, the modeling process itself can become quite time-consuming, even with the ready availability of the IMSL software resources. For example, an analyst employing the IMSL subroutines in a batch processing computer system to perform Box-Jenkins modeling of a time series might perform the following sequence of tasks:

1. Prepare the time series data in the proper format.
2. Plot and visually examine the time series, checking for nonstationarity, trends, seasonality, patterns, etc.
3. Write a program to call the IMSL subroutine that will calculate the mean, variance, autocorrelations and partial autocorrelations of the series.
4. Plot the autocorrelations and partial autocorrelations; this provides much of the information required for identifying the correct model for the series.

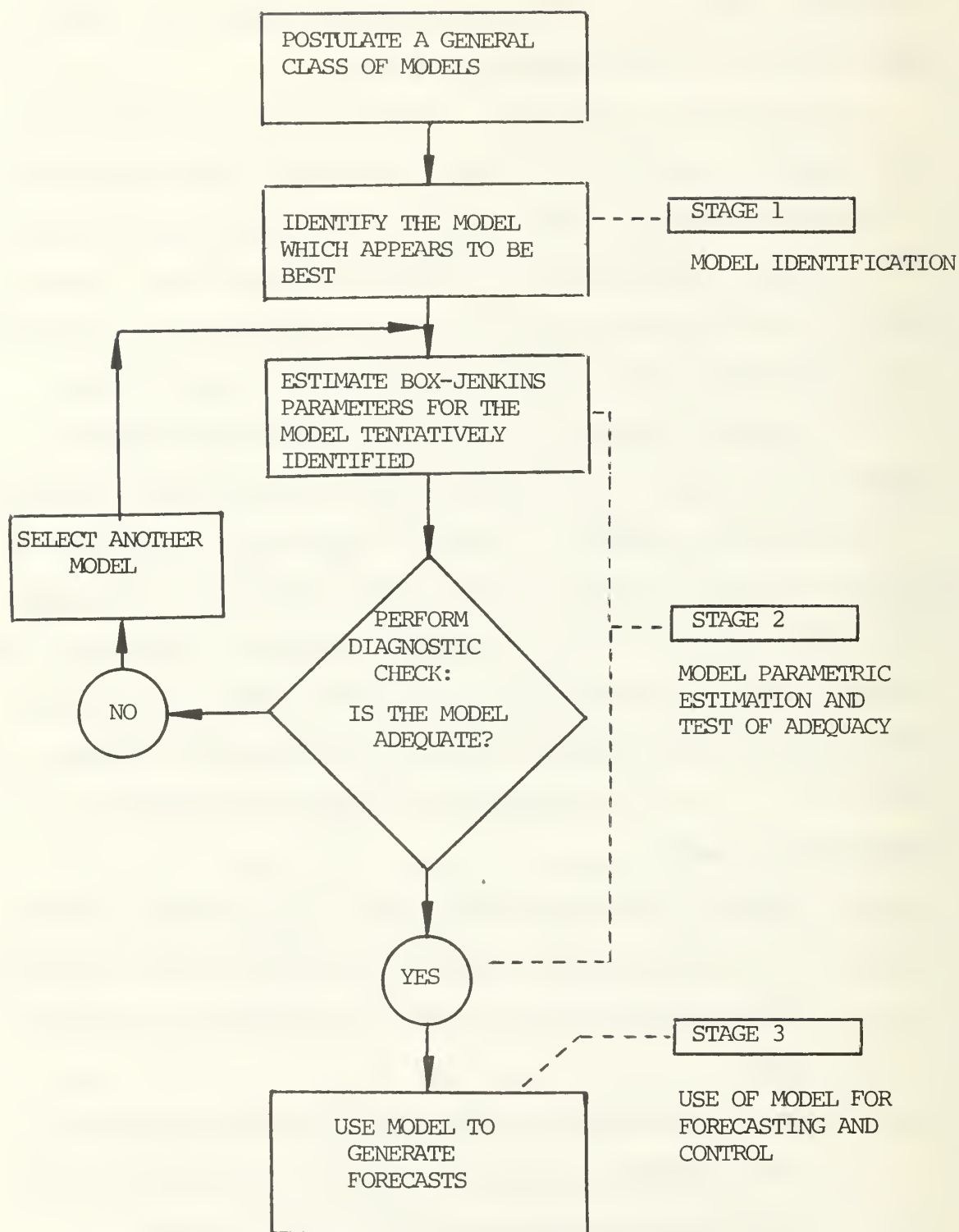


FIGURE 1. The Box-Jenkins Forecasting Methodology

5. Write a program to call the IMSL subroutine which transforms the time series to adjust for seasonal patterns, nonstationary behavior or other behavior which deviates from that assumed for the class of models postulated.
6. Repeat steps (2) through (4) with the transformed data.
7. Review the statistical properties of the autocorrelations and partial autocorrelations for tentative model identification.
8. Write a program to call the IMSL program that estimates the model parameters and computes the residuals.
9. Write a program to perform goodness-of-fit tests for the model.
10. Analyze the model residuals, using steps (1) through (9), as with the original time series.
11. Refine the model using information gained by examination for any structure in the residuals.
12. Repeat the preceding process until no structure remains in the residuals.
13. When an adequate model has been obtained, write a program to call the IMSL subroutine which forecasts and determines confidence intervals for future values of the time series.

Between each successive pair of steps, the analyst must manually intervene in the analysis process, making a decision

subjectively based on the information then available. Thus, much analyst interaction is necessary in order to determine a suitable mathematical model and forecast equation. Even with rapid computer job turnaround time, the process described above could easily consume a full working day or more.

This paper describes an effort to alleviate some of the organizational problems inherent in the modeling of time series using the Box-Jenkins iterative approach. This effort consists of an interactive computer program package which provides access in a structured way to the most useful IMSL and other subroutines necessary for Box-Jenkins time series analysis. This package, called the Time Series Editor, is written for use with the Naval Postgraduate School's Control Program/Cambridge Monitor System (CP/CMS). Since all working programs with the exception of the executive routine itself are written in FORTRAN, the Time Series Editor should be adaptable for use on other FORTRAN-capable time-sharing systems. The Time Series Editor assists the analyst in data preparation and entry, model construction and diagnostic testing, and time series forecasting. In fact, with the User's Guide provided as an Appendix to this report, a complete Box-Jenkins time series analysis can be performed in a short time by even a novice computer user. For this reason, the Time Series Editor could be a valuable instructional tool for laboratory use in a time series course.

A brief overview of Box-Jenkins methodology is provided in Chapter II, to serve as a point of reference for the material which follows. Chapter III contains a description of the algorithm employed to estimate non-linear least-squares parameters for a generalized (seasonal, nonseasonal, stationary, or nonstationary) Box-Jenkins model. Chapter IV contains descriptions of each of the major programs in the Time Series Editor. Chapters V and VI contain examples of actual use of the Time Series Editor, using a non-seasonal and a seasonal series, respectively. Chapter VII includes a summary of the report, and recommendations for future additions to the Time Series Editor. The Appendices include a User's Guide to the Time Series Editor, with instructions, sample user sessions and sample outputs.

II. BOX-JENKINS METHODOLOGY: A PRIMER

This chapter provides a brief overview of the methodology developed for the analysis of time series data by G.E.P. Box and G.M. Jenkins, in order to facilitate understanding of the models and programs employed in the Time Series Editor. For more detail on the material covered in this chapter, the interested reader is referred to the texts by Anderson [Ref. 1], Box and Jenkins [Ref. 4], Pindyck and Rubinfeld [Ref. 15], Nelson [Ref. 14], and Mabert [Ref. 12].

A. LISTING OF NOTATION

The following is a listing of the notation employed in this chapter and in the remainder of this report. Each symbol is provided with a brief title or explanation of its use.

$\{z_t\}$	a time series
N	the length of the time series (number of terms)
$\bar{z}, \hat{\sigma}_z^2$	estimates of the mean and variance of the time series $\{z_t\}$
$\{z'_t\}$	transform of $\{z_t\}$, where $z'_t = (z_t + \zeta)^\lambda$ or $z'_t = \ln(z_t + \zeta)$
$\{W_t\}$	differenced series from $\{z_t\}$, where $W_t = \nabla^d \nabla_s^D z'_t$
ρ_k	autocorrelation of lag k
r_k	estimate of autocorrelation at lag k
∇^d	backward difference operator of order d

∇_s^D	seasonal backward difference operator of order D with season s
B^s	backward shift operator of order s
s	length of the time series' season
ϕ	non-seasonal autoregressive parameters (estimates, $\hat{\phi}$)
Φ	seasonal autoregressive parameters (estimates, $\hat{\Phi}$)
θ	non-seasonal moving average parameters (estimates, $\hat{\theta}$)
Θ	seasonal moving average parameters (estimates, $\hat{\Theta}$)
$\{a_t\}$	series of white noise terms of the process
σ_a^2	variance of the white noise process
$S(\hat{\phi}, \hat{\theta})$	sum of squared residuals of the model
p	number of non-seasonal autoregressive parameters
P	number of seasonal autoregressive parameters
q	number of non-seasonal moving average parameters
Q	number of seasonal moving average parameters
$\psi(B)$	transfer function of a linear filter
ϕ_{kk}	partial autocorrelation of lag k

B. TIME SERIES, STOCHASTIC MODELS AND STATIONARITY

A distinguishing feature of modern time series analysis is that the sequence of observations of a given variable is considered to be a realization of jointly distributed

random variables, such that the time series itself can be viewed as a stochastic process. A discrete stochastic time series, then, is a set of observations (z_1, z_2, \dots, z_t) generated sequentially in time by a set of jointly distributed random variables, where the actual data (z_1, z_2, \dots, z_t) represents a particular realization of some joint probability distribution function $f(z_1, z_2, \dots, z_t)$. With this function determined in some way, a forecast z_{t+k} can be thought of as having been generated by a conditional probability distribution $f(z_{t+k} | z_1, z_2, \dots, z_t)$. Now, the stochastic process which generates the time series (z_1, z_2, \dots, z_t) is said to be stationary if its properties are unaffected by a change in time origin; that is, the process has reached a particular state of statistical equilibrium. This implies that the joint probability distribution associated with the m observations $(z_{t_1}, z_{t_2}, \dots, z_{t_m})$, made at any set of times (t_1, t_2, \dots, t_m) , is identical to that distribution associated with the m observations $(z_{t_1+k}, z_{t_2+k}, \dots, z_{t_m+k})$ made at times $(t_1+k, t_2+k, \dots, t_m+k)$.

For stationary discrete time series, the marginal probability distribution $p(z_t)$ is the same for all times t . Hence, the stochastic process embodied in the time series has a constant mean

$$\mu = E[z_t] = \int_{-\infty}^{\infty} zp(z)dz ,$$

which defines the level about which it fluctuates in time, and a constant variance

$$\sigma_z^2 = E[(z_t - \mu)^2] = \int_{-\infty}^{\infty} (z - \mu)^2 p(z) dz ,$$

which measures its variability about the level of the mean. Since the probability distribution $p(z)$ is the same for all times t , its form can be investigated by plotting a histogram or relative frequency plot of the set of observations (z_1, z_2, \dots, z_t) . Additionally, the mean and variance of the stochastic process can be estimated by the sample averages taken over time:

$$\bar{z} = \frac{1}{N} \sum_{t=1}^N z_t$$

and

$$\hat{\sigma}_z^2 = \frac{1}{N} \sum_{t=1}^N (z_t - \bar{z})^2 .$$

C. THE BASICS OF DIFFERENCE OPERATORS AND LINEAR FILTER MODELS

This section provides a brief discussion of the difference operators currently employed in writing Box-Jenkins models. Those most commonly used are described below in Table I.

Operator	Name	Definition
B^m	Backward Shift Operator of order m	$Bz_t = z_{t-1}$ $B^s z_t = z_{t-s}$
F^m	Forward Shift Operator of order m	$F = B^{-1}$ $Fz_t = z_{t+1}$ $F^m z_t = z_{t+m}$
∇_s^D	Backward Difference Operator of order D and season s	$\nabla^D = (1 - B)^D$ $\nabla_s^D = (1 - B^s)^D$ $\nabla z_t = z_t - z_{t-1}$

Table I. Basic difference operators.

It is noted here that $\nabla^D z_t$ could be written as a binomial expansion; that is,

$$\begin{aligned}
 \nabla^D z_t &= (1 - B)^D z_t = \left(\sum_{j=0}^D \binom{D}{j} (1)^j (-B)^{D-j} \right) z_t \\
 &= \sum_{j=0}^D \binom{D}{j} (-1)^{D-j} z_{t-D+j} .
 \end{aligned}$$

The stochastic models employed in the Box-Jenkins method are based on the idea that a time series in which successive

values are highly dependent can be regarded as having been generated from a series of independent "shocks" a_t . These shock terms, sometimes referred to as white noise terms, are random drawings from some fixed distribution, generally assumed Normal with mean zero and variance σ_a^2 . The white noise process a_t is supposed transformed to the process z_t by what is known as a linear filter, shown in Figure 2 below [Ref. 4, p. 8]. The linear filtering operation takes a

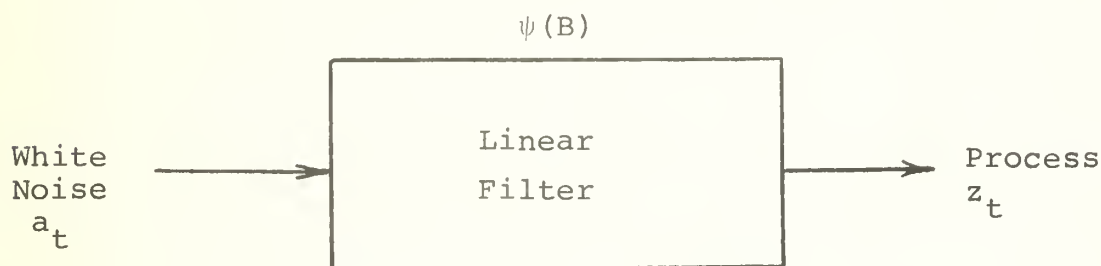


Figure 2. Representation of a time series as the output from a linear filter.

weighted sum of previous observations, so that z_t can be written

$$\begin{aligned}
 z_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \\
 &= \mu + \psi(B) a_t .
 \end{aligned}$$

In general, μ is the parameter that determines the "level" of the process, and

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

is the linear operator that transforms a_t into z_t , called the transfer function of the filter. The sequence ψ_1, ψ_2, \dots formed by the weights may be (theoretically) finite or infinite; if this sequence is finite, or infinite and converges, the filter is said to be stable and the process z_t stationary. The parameter μ is then the value about which the process varies. Otherwise, z_t is called nonstationary, and μ has no specific meaning except as a reference point for the level of the process.

D. HOMOGENEOUS NON-STATIONARY AND SEASONAL PROCESSES

In many cases of interest to the time series analyst, the process being examined is not stationary. Instead, the probabilistic structure of the process generating the time series changes with time. For example, there may be a trend or seasonal pattern in the time series. However, if the series can be found to exhibit behavior which is somewhat homogeneous, then the series can frequently be transformed into a stationary series that can be described by Box-Jenkins models.

A series is said to be homogeneous nonstationary of order d if $w_t = \nabla^d z_t$ is a stationary series for some integer d . Here $\nabla = (1 - B)$ denotes the difference operator such that $\nabla^1 a_t = z_t - z_{t-1}$. The fact that a given series is nonstationary can be determined quickly by examination of

the autocorrelation function, to be discussed later. The series to be modeled should be differenced until the resulting series appears stationary or until the procedure appears to be making no improvement.

A time series is called seasonal when it exhibits cyclical behavior over time, showing a regular period. Examples of seasonal processes include monthly rainfall, monthly crop yields, sale volume of bathing suits, livestock production rates and energy consumption. Seasonal patterns are often easy to spot simply by observing a plot of the series, or through knowledge of the process that generated the series. However, in many cases where the variability of the series is large, seasonal patterns will be difficult to distinguish from other fluctuations. Recognition of seasonality in a series is important, since it provides information useful in modeling and forecasting; as will be shown, the autocorrelation function can assist in recognizing seasonality. The Box-Jenkins modeling approach for seasonal nonstationary time series of season length s is to first transform the series using an appropriate seasonal differencing operator to "sweep out" the seasonal effect; that is, take

$$W_t = \nabla_s^D z_t .$$

If there is a trend or other types of non-stationarity present, as well as the seasonal effect, the series can be

differenced again with the ordinary difference operator until stationarity is achieved. This transform would be written

$$W_t = \nabla^d \nabla_s^D z_t$$

The transformed series W_t is then modeled as a stationary series. For seasonal series where complete stationarity can not be achieved, a class of seasonal models is available; these are discussed later in this Chapter.

E. GENERAL CLASSES OF BOX-JENKINS MODELS

This section will provide brief descriptions of several of the most commonly used classes of Box-Jenkins models, including autoregressive (AR) models, moving average (MA) models, mixed and integrated autoregressive-moving average (ARMA/ARIMA) models, and a generalized seasonal ARIMA model.

1. Autoregressive Models

In autoregressive (AR) models, the current value of the series under study is expressed as a linear combination of previous series values that explain the current observation, plus an unexplained random (white noise) term a_t . For a stationary series, or one that has been transformed to stationarity, the deviation from the mean value for each period, $\tilde{z}_t = z_t - \mu$, can be modeled as dependent on weighted values of the previous deviations from the mean. The AR model can be written

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t .$$

In the expressions which follow the tilde will be dropped, but it is to be understood that the mean has already been subtracted from each observation. This equation is called an autoregressive process of order p , where ϕ_j represents a scalar weighting coefficient for the j^{th} previous period. The model can also be written in more compact form as

$$\phi(B) z_t = a_t ,$$

where

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) ,$$

a polynomial in the operator B . Then, $\phi(B)$ represents a general non-seasonal AR operator. The reason these models are called autoregressive can be seen from recalling regression analysis; here the model relates the dependent variable z_t to a set of explanatory variables ($z_{t-1}, z_{t-2}, \dots, z_{t-p}$), which are previous values of the time series. Therefore, the model is autoregressive. For example, when the value of the time series at time t depends only on the value at time $t-1$, the model becomes an autoregressive model of order 1, denoted AR(1). Similarly, if the value in period t depends on the values in periods $t-1$, $t-2$ and $t-3$,

the model is AR(3). The following equations are the mathematical representation of those two examples:

$$\text{AR}(1) : z_t = \phi_1 z_{t-1} + a_t ;$$

$$\text{AR}(3) : z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t .$$

2. Moving Average Models

In moving average (MA) models, the assumption is made that the current value of the time series at period t can be expressed as a linear combination of the previous forecast errors (or, residuals), $a_t = z_t - \hat{z}_t$. The general equation for this model can be written

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} ,$$

which is called a moving average model of order q , where θ_q represents the scalar weighting coefficient for the q^{th} previous period. Like the AR model, the MA model can be written in more compact operator notation as follows:

$$z_t = \theta(B) a_t ,$$

where

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) ,$$

a polynomial in the operator B. The MA model implies that the analyst can gain valuable information for future predictions by consideration of the weighted sum of a number of previous forecast errors or residuals. The following are examples of MA models of order 1 and order 3, MA(1) and MA(3):

$$\text{MA}(1) : z_t = (1 - \theta_1 B) a_t = a_t - \theta_1 a_{t-1} ;$$

$$\begin{aligned} \text{MA}(3) : z_t &= (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t \\ &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} . \end{aligned}$$

3. Mixed and Integrated Autoregressive-Moving Average Models

A natural extension of the autoregressive and moving average models is to construct a combination of the two. Such mixed processes are called autoregressive-moving average (ARMA) processes, of orders p and q, often written as ARMA(p,q). Box and Jenkins [Ref. 4] notes that for many series encountered in practice, the inclusion of both AR and MA terms in a model results in fewer total parameters than would be required for either a pure AR or pure MA process. The ARMA(p,q) model may be written in the usual form as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} ;$$

in the more compact operator notation, the same model is written

$$\phi(B)z_t = \theta(B)a_t .$$

As an example, an ARMA(2,2) model would be written as:

$$(1 - \phi_1 B - \phi_2 B^2)z_t = (1 - \theta_1 B - \theta_2 B^2)a_t ,$$

or expanded, as:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} .$$

Recall that with differencing or other transformation techniques, a stationary series can be obtained from a homogeneous (seasonal or non-seasonal) non-stationary series. When the differencing technique is integrated directly into the ARMA model, the result is an integrated autoregressive-moving average (ARIMA) model. The model is often written ARIMA(p,d,q), where p and q retain their previous meanings, and d represents the order of (non-seasonal) differencing. The model may also be generalized by including an additional constant term θ_0 that will have the effect of adding a deterministic polynomial trend of order d. The value of θ_0 may be found in general by the formula

$$\hat{\theta}_0 = \hat{\mu} G ,$$

where

$$G = [1 - \sum_{i=1}^p \hat{\phi}_i] [1 - \sum_{j=1}^p \hat{\phi}_j] .$$

The general ARIMA(p,d,q) model can be written compactly as

$$(1) \quad \phi(B) \nabla^d z_t = \theta_0 + \theta(B) a_t .$$

An equivalent form for ARIMA(p,d,q) models that is sometimes seen is called the undifferenced form:

$$\Psi(B) z_t = \theta_0 + \theta(B) a_t ,$$

where

$$\Psi(B) = \phi(B) \nabla^d = \phi(B) (1 - B)^d .$$

The "differenced" form, equation (1), is the form usually seen. In the differenced form, the transfer function $\phi(B)$ is assumed to be stable; that is, all the roots of the difference equation $\phi(B) = 0$ are outside the unit circle. Clearly, $\Psi(B)$ is not a transfer function of a stationary series when $d > 0$, since $\Psi(B)$ has d roots on the unit circle.

4. Seasonal ARIMA Models

The general class of ARIMA models previously discussed can be further generalized to allow for the modeling of seasonal models with the addition of appropriate seasonal operators and parameters. The seasonal model is called an ARIMA model of order $(p,d,q) \times (P,D,Q)_s$, where p,d , and q retain their previous meanings, and P, D, Q and s are as defined in Section A of this chapter, page 17. The ARIMA $(p,d,q) \times (P,D,Q)_s$ model can be written as the product of nonseasonal and seasonal operators:

$$\phi(B)\phi(B^s)\nabla_s^D\nabla^d z_t = \theta_0 + \theta(B)\theta(B^s)a_t.$$

As an example, consider the ARIMA $(1,2,1) \times (1,2,1)_{12}$ model:

$$(1-\phi_1 B)(1-\phi_{1,12} B^{12})\nabla_{12}^2\nabla^2 z_t = \theta_0 + (1-\theta_1 B)(1-\theta_{1,12} B^{12})a_t.$$

F. THE AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTIONS

Mathematically, the autocorrelation at lag k , denoted ρ_k , is defined as:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2}.$$

The autocorrelation function of a time series describes the association (mutual dependence) among values of the same

series taken at different time periods, the difference in time being referred to as the lag, denoted k . Autocorrelation of itself implies nothing about a change in one variable causing a change in another. However, the autocorrelations may provide important information about the structure of a data set and its pattern. In a set of completely random data the theoretical autocorrelations among successive values will always be zero, whereas data values of strong seasonal or cyclical behavior will be highly autocorrelated. This ρ_k can be estimated using the time-averaged sample autocorrelation, denoted r_k , defined as:

$$r_k = \frac{\frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\hat{\sigma}_z^2}$$

A plot of the autocorrelation function versus the lag k , called a correlogram, is very useful for the purpose of determining if a process is stationary and for identifying the appropriate model.

Another function useful in identifying the appropriate model for a given series is the partial autocorrelation function, here called the pauto function. Partial autocorrelations are analogous to autocorrelations in that they indicate the relationship of the values of a time series to various time-lagged values of the same series. However, they differ from autocorrelations in that they measure the

strength of the relationship between values of the series of various lags after the effects of other lags have been removed. In effect, they show the relative strength of the relationship that exists for varying time lags. For time series models of the types to be considered, the partial autocorrelation coefficients can be calculated several ways. One method of calculation, described by Box and Jenkins [Ref. 4, Sec. 3.2.5] begins with the following equation, satisfied by the autocorrelation function, where the ϕ_k are autoregressive parameters:

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k,k-1}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \dots, k$$

This leads to what are known as the Yule-Walker equations, which may be written

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Solving these equations for $k = 1, 2, 3, \dots$, successively, we obtain

$$\phi_{11} = \rho_1,$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

$$\phi_{32} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \text{ etc.}$$

In general, for ϕ_{kk} , the determinant in the numerator has the same elements as that in the denominator, but with the last column replaced with ρ_k . The quantity ϕ_{kk} , a function of the lag k , is called the partial autocorrelation function. Other methods for calculating partial autocorrelations include successively fitting autoregressive processes of orders 1, 2, ..., k by least squares and picking off estimates of $\hat{\phi}_{11}$, $\hat{\phi}_{22}$, ..., $\hat{\phi}_{kk}$ of the last coefficient fitted at each stage. Another method, discussed in Appendix A3.2 of Ref. 4 and due to Durbin, generates estimates of an autoregressive process of order $k+1$ recursively from estimates of autoregressive processed orders k and less. It is derived by observing recursive relationships from the Yule-Walker equations. The recursive formulas are:

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k-j+1} ,$$

$j = 1, 2, \dots, k$, and

$$\hat{\phi}_{k+1,k+1} = \frac{r_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} r_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} r_j} .$$

Another way of looking at partial autocorrelations is to consider a time series $\{z_t\}$ and the model

$$\hat{z}_t = b_0 + b_1 z_{t-1} + b_2 z_{t-2} + \dots + b_{k-1} z_{t-(k-1)} ,$$

where the b values are the least squares estimates of the linear regression coefficients (β 's) in the model

$$z_t = \beta_0 + \beta_1 z_{t-1} + \dots + \beta_{k-1} z_{t-(k-1)} + e_t .$$

Let \tilde{z}_t be the residual of z_t after removing the linear effect of $z_{t-1}, \dots, z_{t-k+1}$ from z_t , such that $\tilde{z}_t = z_t - \hat{z}_t$. Then, the partial autocorrelation of lag k , denoted ϕ_{kk} , is defined to be the simple autocorrelation of lag k for the adjusted series $(\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n)$.

The exact expression for partial autocorrelations of a moving average process is complicated, but an approximate duality between autocorrelations of an autoregressive

process and pautos of a moving average process implies that the pautos of a moving average process would have the same general behavior as the autocorrelations of an autoregressive process of the same order. This approximate duality also implies that the pautos of an autoregressive process of order p will have the same general behavior as the autocorrelations of a moving average process of order p . Thus, one function can be examined and used to confirm the other, and may be of use in model identification, discussed in section G of this chapter.

G. COMMENTS ON MODEL SELECTION EMPLOYING AUTOCORRELATIONS AND PARTIAL AUTOCORRELATIONS

The purpose of this section is to provide a brief overview of the methodology for selecting the proper model for a given time series, through both examination or prior knowledge of the data itself and examination of plots of the autocorrelation and partial autocorrelation functions.

A not yet stationary series should first be differenced or transformed until it appears to be stationary; characteristic of the autocorrelation plot (correlogram) of a non-stationary series is a very slow damping out of the autocorrelations. When this property of the correlogram is observed, the analyst should difference the series until no further improvement is reached. For ARMA(p,q) models, the tentative identification of the model class, that is, the determination of p and q , can be done by examining the

sample autocorrelation and sample partial autocorrelation functions of the given time series with the theoretical auto and pauto functions of members of the general linear class. For most stationary time series, an adequate fit can be found in a model with p and q relatively small, say three or less.

The following figure displays examples of autocorrelation and partial autocorrelation versus lag plots for several basic model types, described in Ref. 17.

Figure 3. Examples of Auto and Pauto Plots

Figure 3(a). Auto and Pauto vs. Lag for AR(1) Model

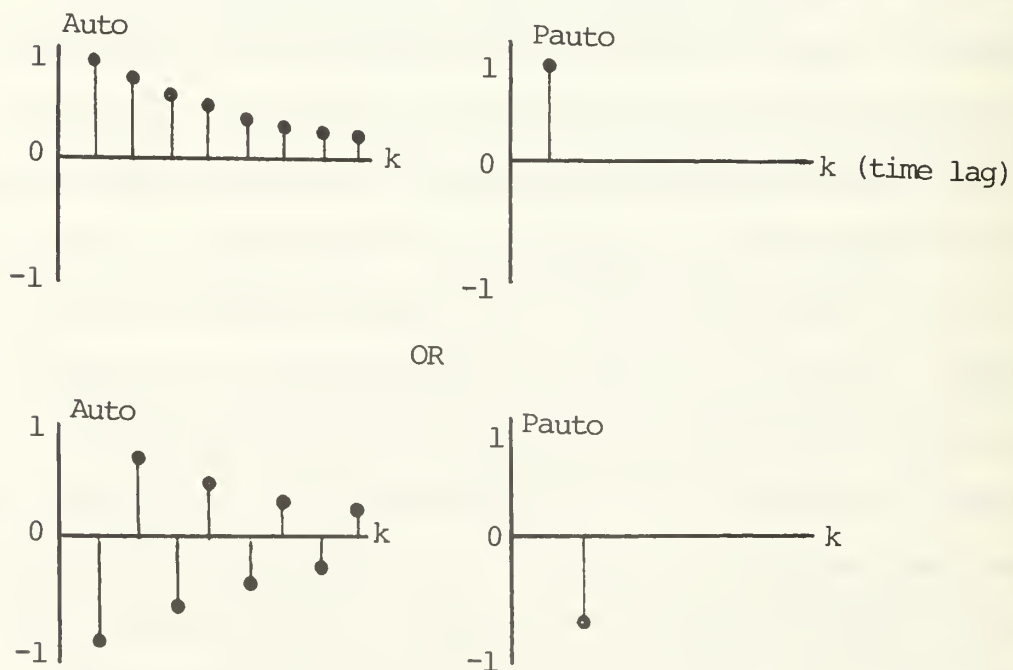


Figure 3(b). Auto and Pauto vs. Lag for AR(2) Model

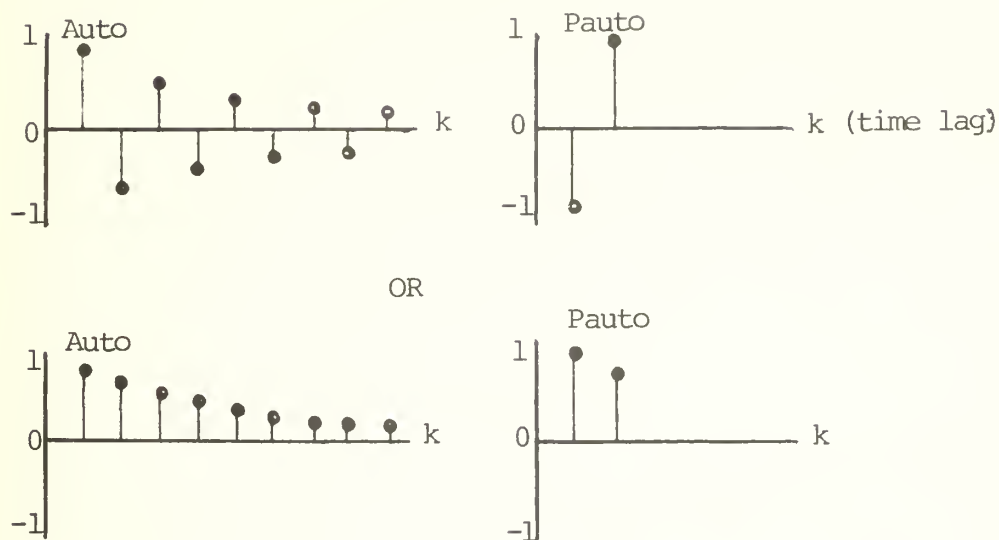


Figure 3(c). Auto and Pauto vs. Lag for MA(1) Model

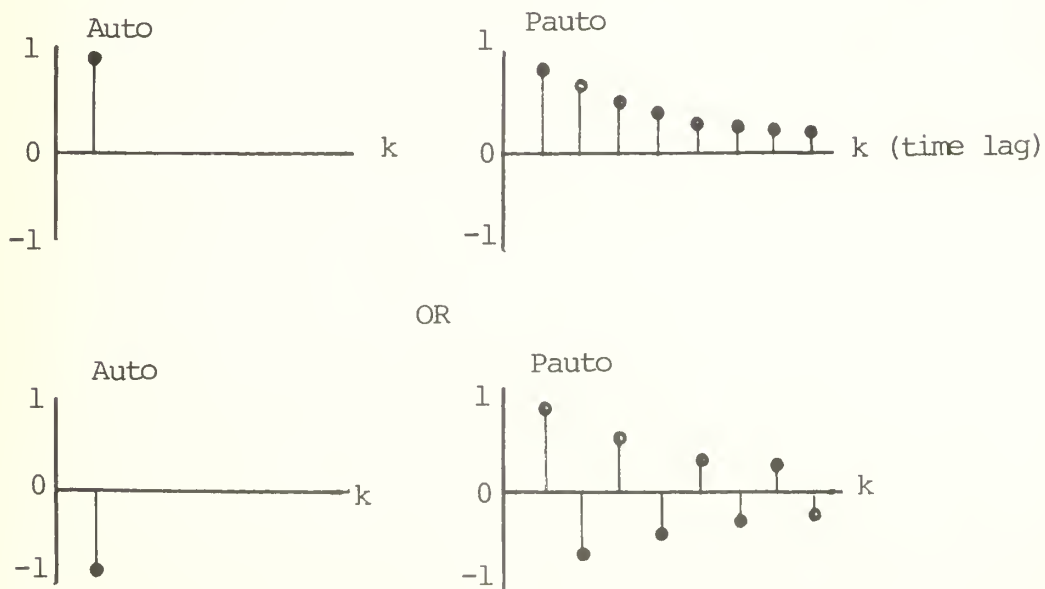


Figure 3(d). Auto and Pauto vs. Lag for MA(2) Model

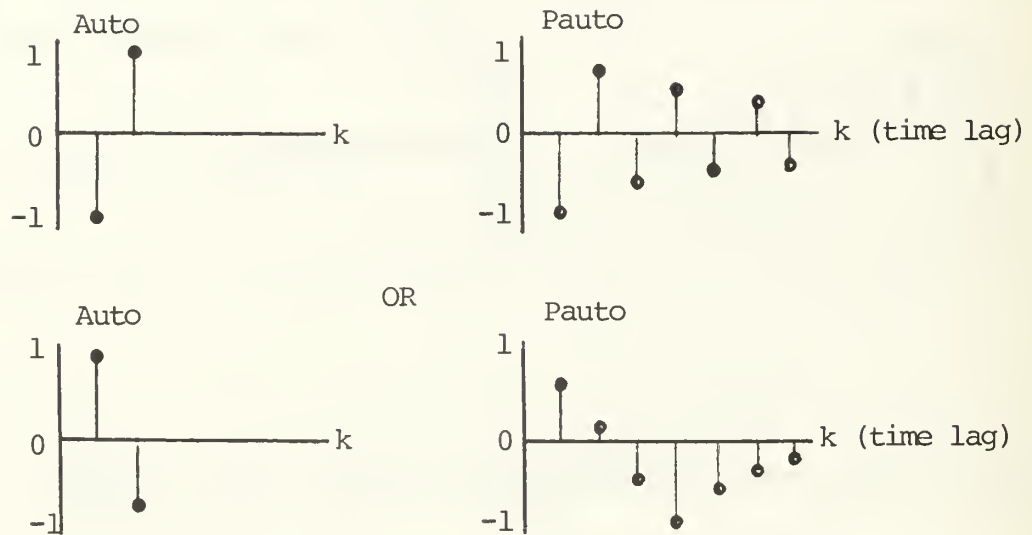
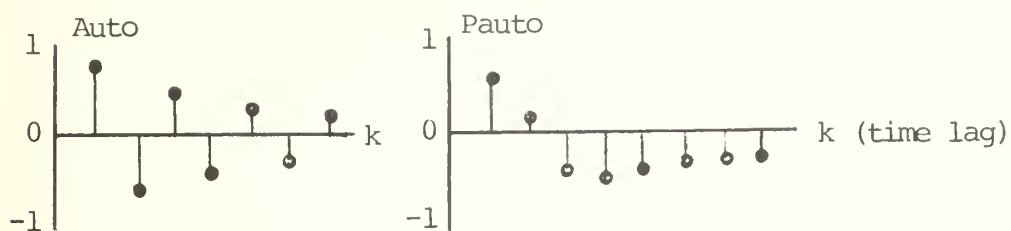
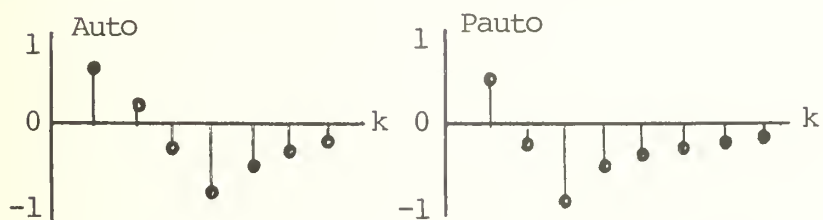


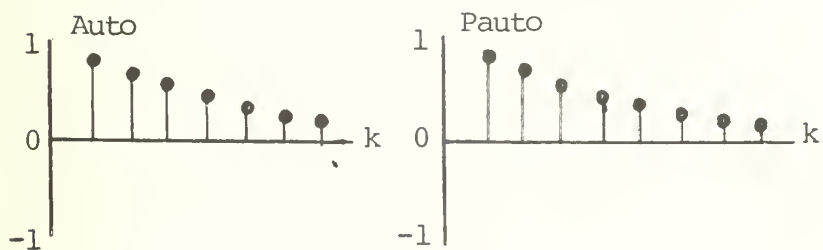
Figure 3(e). Example Auto and Pauto vs. Lag for Mixed ARMA(1,1) Model



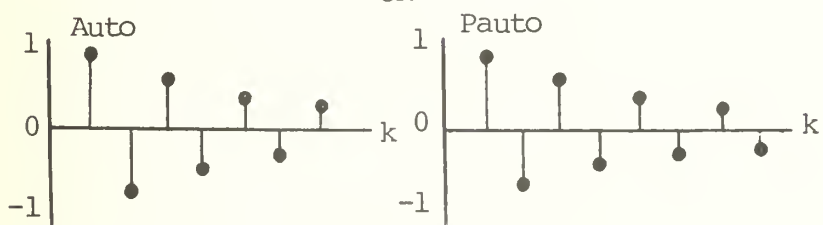
OR



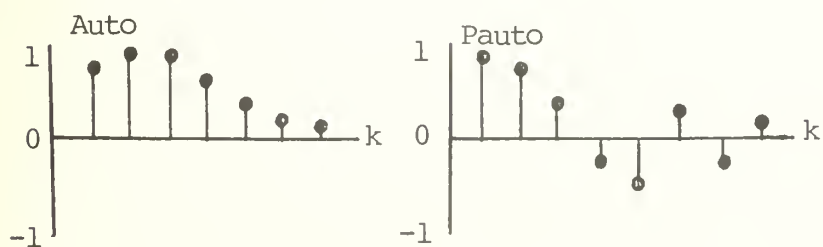
OR



OR



OR



As can be seen from the figures, there is a wide variety of possibilities for appearance of the correlogram and partial autocorrelation plots for even the simplest of models. Sometimes the pattern of estimated autos and pautos can be quickly classified in terms of one of the simpler basic models; however, most real data will generate auto and pauto plots that resemble those of Figure 3(e), those of a mixed model. Here some free association may be required to infer a pattern from the correlogram, or more than one pattern may be implied. However, most of the time a selection of something is possible. As is evident, this is a somewhat subjective process, where the quality of identification will improve with experience.

The following table, adapted from Box and Jenkins [Ref. 4], summarizes the properties of autoregressive, moving average and mixed ARMA processes; an understanding of its contents will be helpful as thumbrules for initial model identification.

It can be noted from examination of Table II that there exists what can be termed a "duality" relationship between the autocorrelations and partial autocorrelations of pure autoregressive and moving average processes. For example, the plot of autocorrelation versus lag for an $AR(p)$ process would appear the same as the plot of partial autocorrelation versus lag for an $MA(q)$ process, for $p = q$. Similarly, the plot of partial autocorrelation versus lag for the $AR(p)$

	Autoregressive Process	Moving Average Process	Mixed Process
Model in terms of previous z_t values	$\phi(B)\tilde{z}_t = a_t$	$\tilde{z}_t = \theta(B)a_t$	$\phi(B)\tilde{z}_t = \theta(B)z_t$
Autocorrelation Function	Infinite and tails off; composed of damped exponentials and/or damped sine waves	finite; that is, there will be q non-zero autocorrelations	Infinite and tails off; composed of damped exponen- tials and/or damped sine waves after the first $q-p$ lags
Partial Autocorrelation Function	finite; that is, there will be p non-zero partial autocorrelations	Infinite and tails off; dominated by damped exponen- tials and/or sine waves	Infinite and tails off; composed of damped exponen- tials and/or sine waves after the first $p-q$ lags

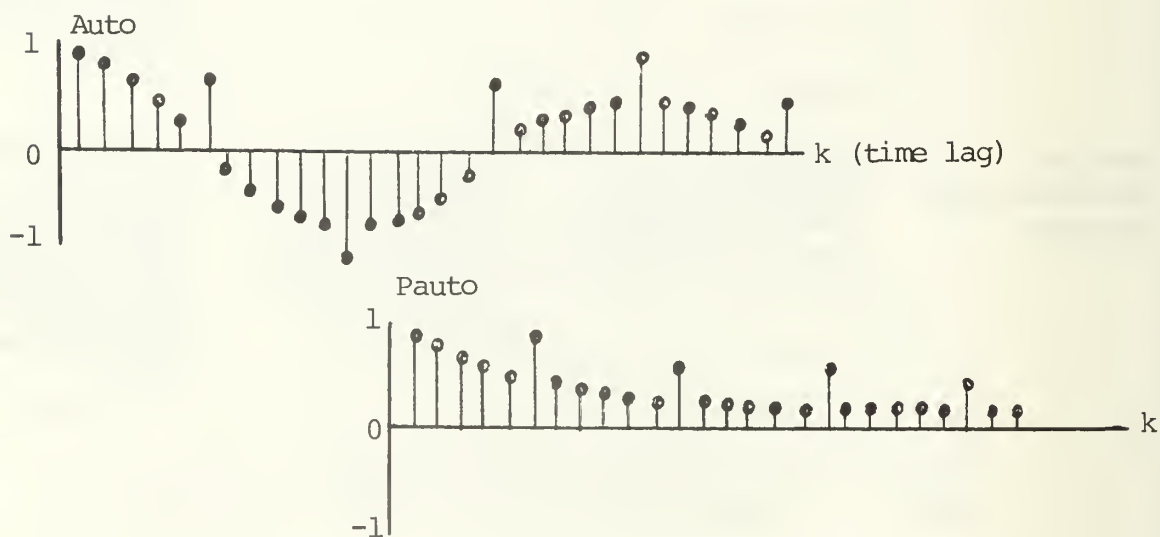
Table II. Properties of Basic ARMA Processes

process would appear the same as the plot of autocorrelation versus lag for an MA(q) process, where $p = q$.

For seasonal series, where differencing has been unable to remove all the nonstationary seasonal behavior, the autocorrelation and partial autocorrelation plots will generally exhibit a "spike" at lags equal to integer multiples of the period of seasonality. For example, if the data are seasonal with a period of 12, then the autocorrelations

$r_{12}, r_{24}, r_{36}, \dots$ and the partial autocorrelations $\phi_{12,12}, \phi_{24,24}, \phi_{36,36}, \dots$ would be amplified. This would indicate a need for one or more seasonal parameters to be included in the model. The following figure is an illustration of such an auto/pauto plot.

Figure 4. Example Auto and Pauto versus lag for mixed ARMA(p,q), with seasonality remaining of period $s = 6$.



H. ESTIMATION OF MODEL PARAMETERS

Suppose a time series has been tentatively identified as an ARIMA(p,d,q) model:

$$\phi(B) \nabla^d z_t = \theta_0 + \theta(B) a_t.$$

There are then $p+q+2$ unknown parameters of the model to be estimated, $(\phi_1, \phi_2, \dots, \phi_p, \theta_0, \theta_1, \theta_2, \dots, \theta_q, \sigma_a^2)$. The

Box-Jenkins procedure separates this estimation into two parts. First, estimates are obtained for the AR and MA parameters $\hat{\phi} = (\phi_1, \phi_2, \dots, \phi_p)$ and $\hat{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$; then estimates are calculated for θ_0 and σ_a^2 , which are functions of $\hat{\phi}$ and $\hat{\theta}$.

The usual procedure is to select those parameter values $\hat{\phi}$ and $\hat{\theta}$ that minimize the sum of squared model errors (residuals). Let $\mu_w = \theta_0 / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$ and $W_t = \nabla^d(z_t - \mu_w)$. Then, it can be shown that $\mu_w = E[W_t]$ and that the model can be rewritten as

$$\phi(B) (W_t - \mu_w) = \theta(B) a_t,$$

or

$$a_t = \theta^{-1}(B) \phi(B) (W_t - \mu_w).$$

Now, let \bar{W} , the sample mean, be the estimate of μ_w (if $d > 0$, then μ_w is usually zero). Also, let

$$\hat{a}_t = \hat{\theta}^{-1}(B) \hat{\phi}(B) (W_t - \bar{W}).$$

Set

$$S(\hat{\phi}, \hat{\theta}) = \sum_{t=1}^N \hat{a}_t^2,$$

where $N = n-d$ is the length of the differenced series. The

objective is to select those parameters $\hat{\underline{\phi}}$ and $\hat{\underline{\theta}}$ such that $S(\hat{\underline{\phi}}, \hat{\underline{\theta}})$ is minimized over all values of the $(p+q)$ -dimensioned parameter space. Since the equation in S is nonlinear in the parameters, iterative search methods are generally used to determine optimal model parameters. Having found these optimal parameters and the minimum S , σ_a^2 can be estimated using

$$\hat{\sigma}_a^2 = S(\hat{\underline{\phi}}, \hat{\underline{\theta}}) / N - p - q .$$

Chapter III contains a more detailed description of a nonlinear algorithm to estimate parameters of a generalized seasonal or nonseasonal ARIMA model.

I. DIAGNOSTIC CHECKING OF MODELS

After the model has been tentatively identified and parameter estimates have been calculated, the next task is to test whether or not the original model specification was correct and that the model itself is adequate in forecasting power. The process of testing the model can take many forms, but will usually include at least the following two steps:

1. Generate a simulated series from the estimated model and compare the simulated series and its auto and pauto functions with the original series and its respective auto and pauto functions. This comparison is essentially subjective.

2. Calculate the residuals of the estimated model, the \hat{a}_t 's, and compare the properties of the residuals with those assumed for the shock terms of the actual process. The residuals should be normally distributed and uncorrelated with each other; that is, there should be no discernable structure in the residuals. There are many quantitative statistical tests and data analytic tools that can be applied to the residuals to test hypotheses of normality and zero autocorrelation.

A plot of the autocorrelation and partial autocorrelation functions of the residuals will provide not only a test of whether or not the residuals are uncorrelated, but, if they are correlated, the plots can be used to suggest improvements to the model. For example, suppose the model was tentatively specified as the ARMA(1,1) model below:

$$(1 - 0.5B)(z_t - 2) = (1 + 0.7B) a_t .$$

Suppose also that the autos and pautos of the model residuals suggested the model $(1 - 0.3B) a_t = u_t$, where the u_t 's are white noise, uncorrelated with variance σ_u^2 . Then, these two models can be combined into the ARMA(2,1) model:

$$(1 - 0.3B)(1 - 0.5B)(z_t - 2) = (1 + 0.7B) u_t .$$

This should be a refinement over the original model.

J. FORECASTING

The objective in forecasting is to predict future values of the time series with as little error as possible. The criterion most often used for forecasting is to calculate that forecast which minimizes the expected mean square forecast error. Therefore, if $\hat{z}_t(\ell)$ denotes the forecast for lead time ℓ from time origin t of the value $z_{t+\ell}$, the objective is to find $\hat{z}_t(\ell)$ such that the objective function

$$E[(z_{t+\ell} - \hat{z}_t(\ell))^2]$$

is a minimum. This forecast is given by taking $\hat{z}_t(\ell)$ as the conditional expectation of $z_{t+\ell}$, given z_1, \dots, z_t :

$$\hat{z}_t(\ell) = E[z_{t+\ell} | z_t, z_{t-1}, \dots, z_1] .$$

The forecast can be easily generated recursively from the mathematical model, utilizing the facts that

$$E[z_{t-j} | z_t, \dots, z_1] = z_{t-j} ,$$

for $j = 0, 1, 2, \dots$ (where t is the current time), and

$$E[z_{t+j} | z_t, \dots, z_1] = \hat{z}_t(j), \text{ for } j = 1, 2, \dots$$

and

$$E[a_t] = \begin{cases} 0 & \text{for times } > t \\ a_t & \text{for times } \leq t \end{cases}$$

For example, consider the Box-Jenkins model

$$(1 - 0.5B + 0.6B^2) z_t = (1 + 0.3B) a_t$$

which could be written in the expanded form as

$$z_t = 0.5z_{t-1} - 0.6z_{t-2} + a_t + 0.3 a_{t-1} .$$

Now, assume it is known that z_{100} is 1.4, z_{99} is 1.0, and the calculation has been made to obtain the residual $\hat{a}_{100} = 0.2$. Then the forecasts of z_{101} , z_{102} and z_{103} made from an origin of $t = 100$ can be calculated from the model as follows:

first, for $\ell = 1$, set

$$\hat{z}_{100}(1) = E[z_{101} | z_{100}, z_{99}, \dots, z_1] ,$$

$$\hat{z}_{100}(1) = E[0.5z_{100} - 0.6z_{99} + \hat{a}_{101} + 0.3\hat{a}_{100}] ,$$

and

$$\hat{z}_{100}^{(1)} = (0.5)(1.4) - (0.6)(1.0) + 0 + (0.3)(0.2) ;$$

then our forecast

$$\hat{z}_{100}^{(1)} = 0.7 - 0.6 + 0.06 = 0.16 ;$$

Next, for $\ell = 2$, set

$$\hat{z}_{100}^{(2)} = E[z_{102} | \hat{z}_{101}, z_{100}, z_{99}, \dots, z_1] ,$$

$$\hat{z}_{100}^{(2)} = E[0.5\hat{z}_{101} - 0.6z_{100} + a_{102} + 0.3a_{101}$$

$$| z_{100}, z_{99}, \dots, z_1] ,$$

$$\hat{z}_{100}^{(2)} = (0.5)(\hat{z}_{100}^{(1)}) - (0.6)(z_{100}) + 0 + 0 ,$$

$$\hat{z}_{100}^{(2)} = (0.5)(0.16) - (0.6)(1.4) = -0.76 ;$$

similarly, for $\ell = 3$, set

$$\hat{z}_{100}^{(3)} = E[z_{103} | \hat{z}_{102}, \hat{z}_{101}, z_{100}, z_{99}, \dots, z_1] ,$$

$$\hat{z}_{100}^{(3)} = E[0.5\hat{z}_{102} - 0.6\hat{z}_{101} + a_{103} + 0.3a_{102}$$

$$| z_{100}, z_{99}, \dots, z_1]$$

$$\hat{z}_{100}^{(3)} = (0.5)(\hat{z}_{100}^{(2)}) - (0.6)(\hat{z}_{100}^{(1)}) + 0 + 0 ,$$

and then

$$\hat{z}_{100}(3) = (0.5)(-0.76) - (0.6)(0.16) = -0.48 .$$

Clearly, this process can be continued into the future as long as desired, recognizing that the expected forecast accuracy will decrease as ℓ increases.

Let

$$e_t(\ell) = (z_{t+\ell} - \hat{z}_t(\ell))$$

denote the forecast error for lead time ℓ beyond the forecast time origin t . It can be shown that $e_t(\ell)$ is given by the relation

$$e_t(\ell) = a_{t+\ell} + \psi_1 a_{t+\ell-1} + \dots + \psi_{\ell-1} a_{t+1} ,$$

where the scalar weights ψ_j are determined from the equation

$$\psi(B) = \hat{\phi}^{-1}(B) (1 - B)^{-d} \hat{\theta}(B)$$

using the known parameter values. The variance of the forecast error is given by

$$E[(e_t^2(\ell))] = (1 + \psi_1^2 + \dots + \psi_{\ell-1}^2)(\sigma_a^2) = \{1 + \sum_{j=1}^{\ell-1} \psi_j^2\} \sigma_a^2 .$$

From these relationships, a confidence interval of n standard deviations about a forecast of lead time ℓ would be given by:

$$c_n(\hat{z}_t(\ell)) = \hat{z}_t(\ell) \pm n[1 + \sum_{j=1}^{\ell-1} \psi_j^2]^{1/2} \hat{\sigma}_a .$$

The value of n could also be thought of as the percentage point for the desired level of confidence using the Normal distribution [Ref. 4]. It can be noted from the definition of $e_t(\ell)$ above that the one-step-ahead forecast error, $e_t(1)$ is simply the value a_{t+1} ; that is,

$$z_{t+1} - \hat{z}_t(1) = e_t(1) = a_{t+1} .$$

This is an explanation for the use of the term "residual" to refer to the white noise or shock terms. Additionally, from the foregoing it is evident that the forecast error variance is a non-decreasing function of the length of the forecast lead time ℓ ; therefore, the confidence bands must become wider as the forecast lead time increases.

III. GENERALIZED ARIMA MODEL PARAMETER ESTIMATION

The purpose of this chapter is to describe the computational details of the algorithm WMARQRDT, employed with appropriate subroutines to estimate Box-Jenkins parameters of a generalized seasonal (or non-seasonal) ARIMA model. This generalized model can be written parsimoniously as follows:

$$\phi(B)\phi(B^S)W_t = \theta_0 + \theta(B)\theta(B^S)a_t ,$$

where

$$W_t \text{ is } \nabla^d \nabla_d^D z'_t ,$$

and

z'_t is the transformed time series (for example,

$$z'_t = \ln z_t \quad \text{or} \quad z'_t = \sqrt{z_t}) .$$

The model employs a Marquardt-type [see Ref. 13] non-linear least squares algorithm for determination of the model parameters, searching in the parameter space for the set of parameters which minimizes the sum of the squared model residuals.

A. MATHEMATICAL DISCUSSION OF THE MARQUARDT ALGORITHM

Reference 13 describes Marquardt's methodology for development of a general algorithm for least-squares estimation of nonlinear parameters. Most algorithms for least-squares estimation of nonlinear parameters have been centered about either a pure linear iterative model based on a Taylor series expansion or using some form of the method of steepest descent (maximum negative gradient). Since both methods have severe potential pitfalls (Taylor series due to divergence of the iterates due to an unfortunate choice of initial values, the steepest descent due to slow convergence after the first few iterations), a method termed "Maximum neighborhood" method was developed by Marquardt which, in effect, performs an optimum interpolation between the Taylor series method and the method of steepest descent, the interpolation being based upon the maximum neighborhood in which the truncated Taylor series gives an adequate representation of the nonlinear model.

As discussed in Ref. 14, the idea of Marquardt's method can be briefly explained as follows. Assume the algorithm begins with a vector of initial parameters $(\phi_0, \phi_0, \theta_0, \theta_0)$. If the method of steepest descent is applied, a certain vector direction h_g , where the subscript g represents the gradient (which gives the direction in which the rate of change is the greatest), is obtained for movement away from the initial point values in the parameter space. Due to potential nonlinearities in the sum-of-squared-residuals

surface, $S(\phi_o, \phi_o, \theta_o, \theta_o)$, in the parameter space, h_g may be the best local direction in which to move toward optimality (minimal sum of squared model residuals), but may not be the best direction. However, the best direction must be within 90 degrees [Ref. 8] of h_g , or else the value of $S(\phi, \phi, \theta, \theta)$ will increase locally. Now, the linearization, or Taylor series method, may result in a different correction vector, h_t . Marquardt found experimentally that for a number of practical problems, the angle, say ξ , between the vectors h_g and h_t , fell in the range $80 \text{ degrees} < \xi < 90 \text{ degrees}$. In other words, the two directions were nearly normal in most cases. The Marquardt algorithm then, provides a method for interpolation between the vectors h_g and h_t , employing a suitable step size, which, in general, converges more quickly than either of the two "parent" methods alone. The remainder of this section describes the mechanics of the algorithm, as applied to the estimation of Box-Jenkins parameters.

B. MODEL INPUT INFORMATION

The following is the required input information to the WMARQRDT parameter estimation model:

$\{W'_t\}$	the transformed time series (for example, $W'_t = \ln z_t$, or $W'_t = \sqrt{z_t}$)
N	the length of the time series
s	the length of the seasonal period (for no seasonality, $s = 1$)

P	the number of seasonal AR parameters, $P \geq 0$
p	the number of non-seasonal AR parameters, $p \geq 0$
Q	the number of seasonal MA parameters, $Q \geq 0$
q	the number of non-seasonal MA parameters, $q \geq 0$
NDIFNS	the number of non-seasonal differences to be taken
NDIFS	the number of seasonal differences to be taken
$\hat{\phi} = (\phi_{1,s}, \dots, \phi_{P,s})$	initial estimates of seasonal AR parameters
$\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)$	initial estimates of non-seasonal AR parameters
$\hat{\theta} = (\hat{\theta}_{1,s}, \hat{\theta}_{2,s}, \dots, \hat{\theta}_{Q,s})$	initial estimates of seasonal MA parameters
$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)$	initial estimate of non-seasonal MA parameters

C. ALGORITHM DESCRIPTION

Initially, the model accepts inputs of the time series length and the (transformed, but undifferenced) time series itself from the user's disc, where it is stored in File FT02F001. The remainder of the input data is then entered interactively on the terminal as prompted by the model. The model then forms the initial parameter estimates into vectors of lengths p, P, q, and Q, and employs them to calculate an initial model residual sum of squares.

Prior to initiating the sum of squares calculation process, the model unravels the polynomial of AR and MA operators, and forms a vector of coefficients for both the AR and MA sides of the model. For example, consider the relatively simple $(2,1) \times (2,1)_{12}$ ARIMA model, written

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi_{1,12} B^{12}) W_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_{1,12} B^{12}) a_t.$$

Unraveling, the model becomes first:

$$(1 - \phi_1 B - \phi_2 B^2)(W_t - \phi_{1,12} W_{t-12}) = (1 - \theta_1 B - \theta_2 B^2)(a_t - \theta_{1,12} a_{t-12}),$$

and then takes the form:

$$\begin{aligned} W_t - \phi_{1,12} W_{t-12} - \phi_1 W_{t-1} + \phi_1 \phi_{1,12} W_{t-13} - \phi_2 W_{t-2} + \phi_2 \phi_{1,12} W_{t-14} \\ = z_t^{-\theta_{1,12}} a_{t-12} - \theta_1 a_{t-1} + \theta_1 \theta_{1,12} a_{t-13} - \theta_2 a_{t-2} + \theta_2 \theta_{1,12} a_{t-14}. \end{aligned}$$

This is represented in the computer as four vectors: two vectors, π and γ , composed of the expansion AR and MA coefficients respectively, and two vectors composed of the indices of the time series terms associated with the coefficients. For the model in the example above, these vectors would be written as follows:

$$\pi = (1, \phi_{1,12}, \phi_1, (\phi_1)(\phi_{1,12}), \phi_2, (\phi_2)(\phi_{1,12}))$$

$$\text{INDEX}_{\text{AR}} = (0, 12, 1, 13, 2, 14)$$

$$\gamma = (1, \theta_{1,12}, \theta_1, (\theta_1)(\theta_{1,12}), \theta_2, (\theta_2)(\theta_{1,12}))$$

$$\text{INDEX}_{\text{MA}} = (0, 12, 1, 13, 2, 14).$$

This format for unraveling of a generalized product of polynomials of seasonal and non-seasonal parameters permits simplified calculations in the determination of the sum of squared residuals, since the arithmetic need only be performed for the non-zero parameter values, those contained in the π and γ vectors. In the actual calculations, performed in FORTRAN, the index vector origin is translated to the right by the amount $p + P_s + q + Q_s$, called IADDIT in the model, in order to prevent obtaining zero or negative subscripts.

The calculation of the residual sum of squares is a three stage process. First, the model calculates ten values of e_t , the "forward" white noise terms, using the following relationship:

$$e_t = W_t - \sum_{i=2}^{p+P_s} \pi_i W_{t-i} + \sum_{j=2}^{q+Q_s} \gamma_j e_{t+j}$$

The assumption is made that $(e_{11}, e_{12}, \dots, e_N)$ are all equal to zero. Having found the desired values for $(e_{10}, e_9, \dots, e_1)$, we can now backcast the necessary values of W_t , in order to enable us to calculate the required white noise terms a_t , as follows:

$$W_t = e_t + \sum_{i=2}^{p+Ps} \pi_i W_{t+i} - \sum_{j=2}^{q+Qs} \gamma_j e_{t+j} ,$$

solving for $(W_0, W_{-1}, W_{-2}, \dots, W_{1-(p+sP+q+sQ)})$. Now, with the W_t 's backcast, we can proceed to calculate the estimated white noise terms (residuals) using this relationship:

$$a_t = W_t - \sum_{i=2}^{p+Ps} \pi_i W_{t-i} + \sum_{j=2}^{q+Qs} \gamma_j a_{t-j} .$$

The values of a_t are calculated for $t = 1, \dots, N$. Once the a_t 's are calculated, it is a simple matter to calculate the sum of squares, using

$$S(\phi, \phi, \theta, \theta) = \sum_{t=1}^N a_t^2 .$$

Having calculated the residual sum of squares for the initial parameter estimates, we form the parameters into a vector $\hat{\beta}$, such that

$$\hat{\beta} = (\beta_1, \beta_2, \dots, \beta_k) ,$$

where $k = p+P+q+Q$; that is,

$$\hat{\beta} = (\phi_1, \dots, \phi_p, \phi_{1,s}, \dots, \phi_{p,s}, \theta_1, \dots, \theta_q, \theta_{d,s}, \dots, \theta_{Q,s}) .$$

The subroutine PARSH then calculates the derivatives

$$x_{i,t} = - \frac{\partial a_t}{\partial \beta_i} ,$$

over all values of $t = 1, \dots, N$ and $i = 1, \dots, k$. Using the residuals $\{a_t\}$ calculated earlier, the derivatives are estimated numerically using a perturbation of each parameter of amount δ , such that

$$x_{i,t} = \{a_t(\beta_1, \dots, \beta_k) - a_t(\beta_1, \dots, \beta_i + \delta, \dots, \beta_k)\} / \delta .$$

Then with $\{a_t\}$ and $x_{i,t}$ supplied for the current parameter values, the following quantities are formed:

1. The $k \times k$ matrix $A = \{A_{ij}\}$, where

$$A_{ij} = \sum_{t=1}^N (x_{i,t})(x_{j,t})$$

2. the vector $G = \{g_i\}$, where

$$g_i = \sum_{t=1}^N (x_{i,t})(a_t)$$

3. the scaling quantities $D_i = \sqrt{A_{ii}}$.

Then, the modified (scaled and constrained) linearized equations

$$[A^*] [h^*] = [g^*]$$

are formed, where

$$A_{ij}^* = A_{ij}/D_i D_j, \quad i \neq j,$$

$$A_{ii}^* = A_{ii}^* + \xi, \quad \xi = .01,$$

and

$$g_i^* = g_i/D_i.$$

It can be noted here that the $[A_{ij}^*]$ matrix is in fact a correlation matrix for the model parameters. Those equations are then solved for h^* , which is scaled back to give the parameter correction vector h_j , where

$$h_j = h_j^*/D_j.$$

Then, the new parameter values are formed from the h vector, such that

$$\beta_{NEW} = \beta_{OLD} + h,$$

and the sum of squared residuals for the β_{NEW} set of parameters is calculated.

Now, if $S(\beta_{\text{NEW}}) < S(\beta_{\text{OLD}})$, the parameter corrections h are tested. If all elements of the h vector are smaller than some epsilon, say .0001, then convergence is assumed and the $k \times k$ matrix $[A^*]^{-1}$ is used to calculate the covariance matrix of the estimates; otherwise, β_{OLD} is reset to β_{NEW} , the constant ξ is reduced by a factor F , say .1, and computation returns to the calculation of a new set of derivatives.

However, if $S(\beta_{\text{NEW}}) > S(\beta_{\text{OLD}})$, the constraint parameter ξ is increased by the factor F , and computation is returned to the stage where the A matrix is formed. In all but unusual cases, a minimal sum of squares will be found. However, an upper bound can be placed on ξ , and when it is exceeded, the search is terminated. When convergence is reached, the desired output information is calculated.

D. MODEL OUTPUT AND DIAGNOSTIC CHECKS

The output includes:

1. $\beta = (\beta_1, \dots, \beta_k)$ the least-squares estimates of

$$\phi_1, \dots, \phi_p$$

$$\phi_1, \dots, \phi_p$$

$$\theta_1, \dots, \theta_q$$

$$\theta_1, \dots, \theta_q$$

2. $\hat{\sigma}_a^2$ the residual variance, $\frac{1}{N-p-q-P-Q} S(\hat{\phi}, \hat{\phi}, \hat{\theta}, \hat{\theta})$

3. V the covariance matrix of the parameter estimates, formed from

$$V = \{V_{ij}\} = (A^T A)^{-1} \hat{\sigma}_a^2$$

4. s_i the standard errors of the parameter estimates,

$$s_i = \sqrt{V_{ii}} \quad , \quad i = 1, \dots, p+q+P+Q$$

5. R_{ij} the correlation matrix, obtained from

$$R_{ij} = V_{ij} / \sqrt{V_{ii} V_{jj}}$$

6. $\hat{\theta}_0$ the overall constant term, where
 $\hat{\theta}_0 = \hat{\mu} G$, for

$$G = (1 - \sum_{i=1}^p \hat{\phi}_i) (1 - \sum_{j=1}^P \hat{\phi}_j) \quad \text{and}$$

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N w_t$$

Finally, diagnostic checks on the model are performed. This includes calculation of residual autocorrelations $\hat{r}_{aa}^{\wedge}(k)$, obtained from these formulas:

$$\hat{r}_{aa}^{\wedge}(k) = \hat{c}_{aa}^{\wedge}(k) / \hat{c}_{aa}^{\wedge}(0) \quad ,$$

where

$$c_{\hat{a}\hat{a}}^{\wedge}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (\hat{a}_t - \bar{a}) (\hat{a}_{t+k} - \bar{a}) ,$$

$$\bar{a} = \frac{1}{N} \sum_{t=1}^N a_t ,$$

and k goes from 1 to the maximum desired lag K , usually 40. Additionally, a chi-square statistic is calculated from

$$\chi_{(v)}^2 = N \sum_{k=1}^K r_{\hat{a}\hat{a}}^2(k) ,$$

and is compared with a chi-square distribution with $v = K - p - q - P - Q$ degrees of freedom.

Examples of input methodology and appearance of output for this model are contained in later chapters of this paper.

IV. DESCRIPTION OF THE TIME SERIES EDITOR

In this chapter, descriptions are given for each of those programs in the Time Series Editor which interact with the user. There are fifteen separate program modules currently included in the Editor; there are other programs utilized by the Editor that are completely transparent to the user, and are not described here. However, listings for all programs contained in the Editor are provided in Appendix D to this report.

No attempt is made in this chapter to describe the actual mathematical calculations or algorithms that are performed by the programs. Rather, the objective is to provide general descriptions of what each program can do for the user, and how the user interacts with the programs. The programs described in this chapter are, in the order presented: TIMESER EXEC, ZFORMAT, CMSWORK, TRANS, DIFF, PLOT, AUTO, ESTIMATE, YESTSEAS, WMARQRDT, XSUMSQ, FORECAST, ROOTS, SIMULATE and GENERATE, and HELP.

A. THE EXECUTIVE PROGRAM

The heart of the Time Series Editor is a master program called TIMESER EXEC which provides file control for all of the other program modules, controls input and output, supervises the necessary CP/CMS protocol and provides instructions interactively to the user concerning the contents of the

Time Series Editor and how each program can be used.

The TIMESER EXEC program is written in the IBM-360 CP/CMS Executive language. It is the only program in the package not written in FORTRAN; consequently, it should be the only program that would need modification if the Editor were to be adapted to another FORTRAN-capable time-sharing system.

After the user has logged into CP/CMS and linked to the disc space containing the Time Series Editor (instructions for this procedure are provided in detail in the User's Guide to the Time Series Editor, included as Appendix A to this report), the entire Editor package is made available to the user by the command TIMESER. On entry of this comand, the executive routine, TIMESER EXEC, can provide a guided tour through the Editor. It will briefly describe what the Editor can provide and asks the user what option he wants to use. On the basis of the user's response, the Editor then advises the user what input data is required and how it should be entered, either through a data file entered offline or parametric data entered interactively at the terminal. When the user selects an option for execution, the TIMESER EXEC routine loads the appropriate program package and automatically manipulates the required input and output data files. For the more experienced user who does not require detailed user instructions, there is a shortened version of the TIMESER EXEC. Upon entry into the shortened version, the user is immediately asked which

option he desires. This version of the Editor is entered by adding the argument "s" when logging into the Editor; that is, type TIMESER S when logging in. The TIMESER S (shortened) version of the Editor provides exactly the same program options to the user as the longer TIMESER version.

B. DATA AND FILE MANAGEMENT

Whenever data is required, the user is prompted at the terminal by either the TIMESER EXEC or the program module being executed. In most cases, the necessary user response is a short alphanumeric character input during execution using the terminal keyboard. However, when the time series data itself is entered, the user may enter the data by card deck offline. Similarly, most output is provided right at the user's terminal. Output such as listings of transformed series or plots are sent to the offline printer to conserve time and provide hard copies of the results. Detailed descriptions of the input and output requirements of each program module are provided in this chapter, and also in abbreviated tabular form in Appendix A, the User's Guide. However, there are some general principles of data management that apply to all programs. These are described in this section.

1. Offline Data Entry

The Time Series Editor requires that the user's time series reside in FILE FT02F001. The data can be read offline

via cards and directed to FILE FT02F001. An example of the required format for the input card deck for the Time Series Editor is provided in Figure 5. It is noted here that the data format expected by all programs is FORMAT(5F15.6), preceded by a card containing the length of the time series in FORMAT(I3).

2. ZFORMAT Program

In many cases, the user will want to analyze time series data that are already available in a format other than that required by the Editor. To spare the user the tedious task of retyping his deck in the required format, the ZFORMAT program was written. This program converts a time series data file (FILE FT03001) written in any FORTRAN format into the proper FORMAT(5F15.6) onto FILE FT02F001. The length of the series is entered as an input on the terminal. The data are then ready to use in the Editor, with the original file left in FILE FT03F001. Figure 6 shows a sample input deck for such non-standard data. Figure 7 gives a sample session where ZFORMAT is used.

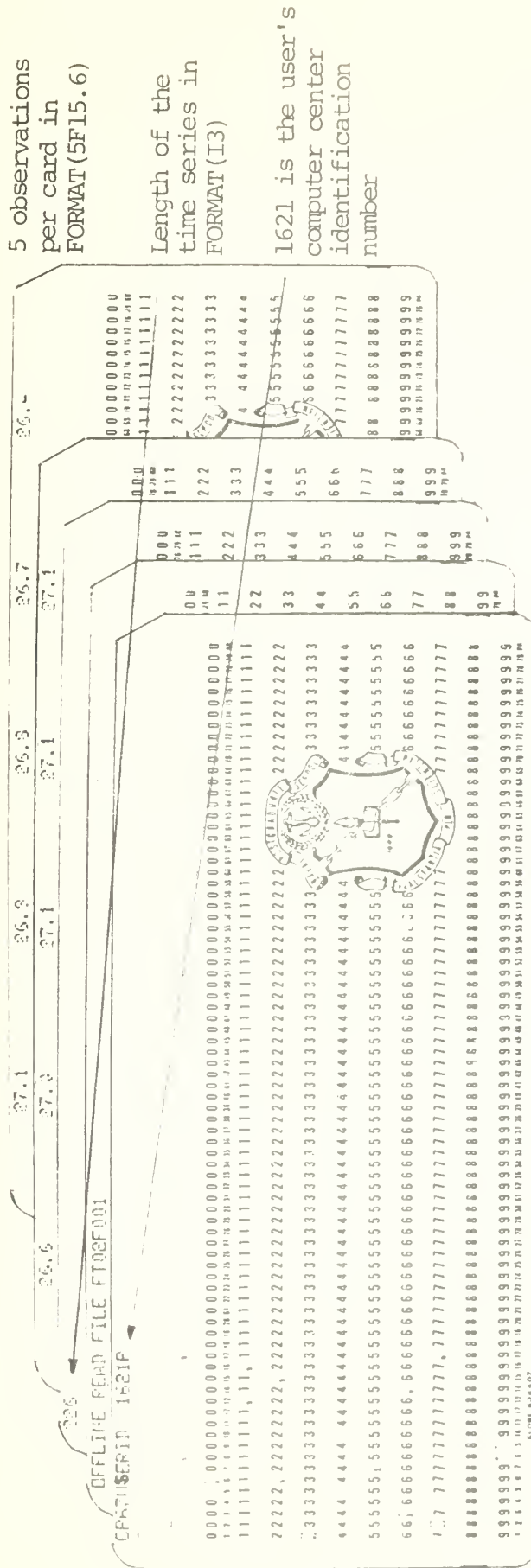


Figure 5. Card deck arrangement for OFFLINE READ of time series input data in FORMAT(5F16.5)

16.3	23.5	32.6	19.8
OFFLINE READ FILE FT03F001			
CPAUSERID 1621P			
55.2	42.3		
0000	0000	0000	0000
1111	1111	1111	1111
2222	2222	2222	2222
3333	3333	3333	3333
4444	4444	4444	4444
5555	5555	5555	5555
6666	6666	6666	6666
7777	7777	7777	7777
8888	8888	8888	8888
9999	9999	9999	9999

Figure 6. Card deck arrangement for OFFLINE READ of time series input data in any FORMAT

timeser s

ENTER LETTER FOR OPTION YOU WANT.

Z

IS YOUR DATA IN FILE FT03F001?

Y

EXECUTION BEGINS...

IS YOUR TIME SERIES DATA NOW IN FILE FT03F001?

Y

ENTER THE LENGTH OF YOUR TIME SERIES VIA FORMAT I3.

010

NOW ENTER THE FORTRAN FORMAT FOR YOUR TIME SERIES DATA,
INCLUDING PARENTHESES; FOR EXAMPLE, TYPE: (5F15.7).

f10.2

YOUR FORMAT IS:

FORMATF10.2

IS THIS CORRECT?

n

NOW ENTER THE FORTRAN FORMAT FOR YOUR TIME SERIES DATA,
INCLUDING PARENTHESES; FOR EXAMPLE, TYPE: (5F15.7).

(f10.2)

YOUR FORMAT IS:

FORMAT(F10.2)

IS THIS CORRECT?

Y

YOUR DATA IS NOW TRANSFORMED INTO THE PROPER FORMAT
FOR USE IN THE TIME SERIES EDITOR, LOCATED IN FILE
FT02F001. YOUR ORIGINAL DATA IS IN FILE FT03F001.

DO YOU WANT TO GO AGAIN?

n

CONTROL RETURNED TO CMS

R;

Figure 7. Sample user session with ZFORMAT program

3. Interactive Keyboard Input

a. Numeric Input

When the user is prompted at the terminal to enter such numerical values as the number of AR parameters, estimates of parameters or the number of differences to be taken, he must enter these values before program execution can continue. The user should enter the data according to standard FORTRAN practice. For example, integer data should be entered (without a decimal point) for counts, such as the length of a seasonal period, and for names beginning with the letters I through N; floating-point data, such as initial estimates of parameters, should be entered with a decimal point. Because a typed decimal point overrides a floating-point format, it is not necessary for the user to concern himself with the format for floating-point input data. However, care must be taken when entering integer data because it must be right-justified in its format field. The user is advised in each case when the format is other than I1. For example, suppose the user desires to perform analysis on a series having only 90 observations. If the program that he is executing requires the length of the time series, the user will receive the following request:

ENTER LENGTH OF THE TIME SERIES VIA FORMAT I3

The user would then enter:

column 123
b90 (where b represents a blank space).

If the blank were omitted, the program would read the series length as 900, and problems in execution would occur.

b. Alphabetic Input

The Editor often asks the user to respond with alphabetic input. Generally, this is in the form of the response to a question requiring a yes or no answer, or a response to the question of which program to execute. The editor has been programmed to read only the first letter of such responses. Therefore, the user need only enter the first letter of the response for each such inquiry. For example, the user should enter Y for yes, N for no, P for the PLOT option, X for the XSUMSQ option, W for the WMARQRDT option, etc. The other alphabetic input is generally entry of titles to plots; these can be any combination of numeric and alphabetic characters, as long as they do not exceed 72 columns in length. For example, a suitable title would be:

PLOT OF AUTOS AND PAUTOS FOR SERIES C DATA (2 ORDINARY DIFFERENCES).

4. Output

Most of the results are written out at the user's terminal. However, in some cases the output is printed

offline to conserve time. Such results as plots and transformed time series are written onto various files (FT03F001, FT09F001, FT08F001), and the TIMESER EXEC program causes them to be automatically printed offline under the user's identification number (USERID). The plots and files can also be printed out at the terminal using the usual CMS commands in the CMS mode, outside the Editor, or in the CMSWORK program (see next section). However, the user should be aware that plotting graphs at the terminal is generally quite a slow process.

5. CMSWORK Program

This program permits the user to enter the CMS environment to perform routine CMS functions without leaving the Time Series Editor. Commonly used CMS commands include file name alteration, file erasure, file offline printing or punching, obtaining a disc status, and listing all disc-resident files. However, essentially all CMS commands can be issued while in CMSWORK. Figure 8 provides an example of the use of the CMSWORK program.

C. TRANS PROGRAM

The TRANS program takes a given time series in File F102F001 and performs a transformation upon each data element. The options include a change of scale, a power transformation, a shift of the origin, and a natural logarithm transformation. The original data is written onto a file

timeser s

ENTER LETTER FOR OPTION YOU WANT.

C

ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

stat

P (191): 30 FILES; 263 REC IN USE, 33 LEFT (of 296), 89% FULL (2 CYL)

listf * ft02f002

FILE NOT FOUND

17.14.13 LISTF * FT02F002

!!! E(00002) !!!

listf * ft02f001

FILENAME FILETYPE MODE NO.REC. DATE

FILENAME	FILETYPE	MODE	NO.REC.	DATE
SERG	FT02F001	P1	3	9/16
SERC	FT02F001	P1	5	9/16
LNSERG	FT02F001	P1	3	9/16
FILE	FT02F001	P1	3	9/16

cp q f

FILES:- NO RDR, NO PRT, NO PUN

erase file ft08f001

offline print file ft02f001

offline print serg ft02f001

offline print serc ft02f001

erase file ft02f001

stat

P (191): 28 FILES; 240 REC IN USE, 56 LEFT (of 296), 81% FULL (2 CYL)

&goto -ques

DO YOU WANT TO GO AGAIN?

n

CONTROL RETURNED TO CMS

R;

Figure 8. Sample user session with CMSWORK program

called DATA FT02F001, while the transformed series is written onto FILE FT02F001. The parameters of the transform itself are written onto FILE FT07F001. The TRANS program is self-contained, using no other programs.

D. DIFF PROGRAM

The DIFF program takes a given time series in FILE FT02F001, and performs non-seasonal and/or seasonal differencing operations on the series in order to help achieve stationarity for proper modeling. The user must indicate if seasonal differences are required; if so, the length of the season is input. The user is then required to enter the number of seasonal and nonseasonal differences. After execution of DIFF, the original data are in FILE FT02F001, and the differenced data are written onto FILE FT03F001. To use the differenced data in any other program, CMSWORK program must be entered and the files altered to save the original data (if desired) and to transfer the differenced data to FILE FT02F001. The CMS commands to perform this shift of files would look like:

```
ALTER FILE FT02F001 P1 SAVEDATA FT02F001 P1
R;
ALTER FILE FT03F001 P1 FILE FT02F001 P1
R;
```

The DIFF program uses the IMSL subroutine called FTRDIF.

E. PLOT PROGRAM

The PLOT program plots any given time series which resides in FILE FT02F001. Other than the time series, which must be on the user's disc or entered offline, the program requires only that an identification title for the plot itself be entered during execution. The plot is automatically printed offline. The PLOT program uses a modified PLOTP subroutine called PLOT8 from the IBM Scientific Subroutine Package Library (SSPLIB).

F. AUTO PROGRAM

The AUTO program takes a given time series in FILE FT02F001 and calculates summary statistics of utility in Box-Jenkins modeling. The statistics include the sample mean, variance, autocorrelations and partial autocorrelations for lags one through 40. Part of those statistics are printed out at the user's terminal, while plots of the autocorrelations and partial autocorrelations versus lag are printed offline. These plots will also be written onto the user's disc in FILE FT08F001, and may be plotted at the terminal in CMS outside the Editor or in the CMSWORK program; recall that this is a time-consuming process. The AUTO program provides the user with essential information about stationarity, seasonality and model identification. With the Box-Jenkins procedures, the second moments (autocorrelations and partial autocorrelations) are valuable tools in tentative model identification. Additionally, AUTO

is useful in the diagnostic checkout phase of model building, when examining the model residuals for structure. The AUTO program uses the IMSL subroutine FTAUTO.

G. ESTIMATE PROGRAM

After the user has tentatively identified a model for a series through analysis of the series plot, autocorrelations and partial autocorrelations, ESTIMATE should be executed to calculate maximum likelihood estimates of the model parameters. ESTIMATE is written to calculate parameters for non-seasonal models only; for seasonal models, the programs YESTSEAS and WMARQRDT should be used for parameter estimation. The ESTIMATE program requires that the series to be modeled (already transformed and differenced to achieve stationarity) reside in FILE FT02F001; additionally, the program will instruct the user to input the number of MA and AR parameters in the model, as well as the number of differences taken. The general model for which ESTIMATE calculates parameters is:

$$\phi(B) W'_t = \theta_0 + \theta(B) a_t$$

where $W'_t = \nabla^d z'_t$ and z'_t is the transformed value of z_t .

ESTIMATE calculates estimates of the AR parameters for the undifferenced form of the model, the estimated MA parameters and MA constant term, the residual variance, autocorrelations and partial autocorrelations of the residuals, plots of

these autocorrelations and partial autocorrelations, and a Chi-square goodness-of-fit statistic for the model. The original series to be modeled remains in FILE FT02F001 and the residuals of the model are stored in FILE FT02F001. ESTIMATE uses the IMSL subroutine FTMAXL (a nonlinear gradient search algorithm) to estimate the model parameters.

H. YESTSEAS PROGRAM

The YESTSEAS program is used to calculate initial estimates of autoregressive and moving average parameters for a generalized seasonal Box-Jenkins model. These initial estimates are used as inputs to the program WMARQRDT, which then refines them to determine the parameter estimates that minimize the sum of the squared residuals. The general model assumed in YESTSEAS is

$$\phi(B) \phi(B^S) \nabla^d \nabla_S^D z_t = \theta(B) \theta(B^S) a_t .$$

The YESTSEAS program requires that the seasonal time series for which initial parameter estimates are required is located in FILE FT02F001; it should be transformed, if desired, but generally not differenced, since YESTSEAS performs the required (seasonal and/or non-seasonal) differencing. Terminal inputs required by the program include the number of seasonal and non-seasonal differences to be taken (if data is already differenced, enter zero for these values), the length of the seasonal period and the number of each

type of parameter to be estimated (that is, the number of each of the seasonal and non-seasonal AR and MA parameters). The output of the model, initial estimates for the requested parameters, is printed out at the terminal. YESTSEAS uses the IMSL subroutines FTRDIF, FTMAXL, FTAUTO and FTARPS, as well as a modified version of FTMAPS.

I. WMARQRDT PROGRAM

The WMARQRDT program uses the initial parameter estimates calculated by YESTSEAS or other means as starting points for calculation of parameters of a generalized seasonal Box-Jenkins model which minimize the sum of the squared residuals. The general seasonal model assumed by WMARQRDT can be written:

$$\phi(B)\phi(B^S)\nabla^d\nabla_s^D z_t = \theta_0 + \theta(B)\theta(B^S)a_t.$$

Details of the Marquardt-type non-linear least squares algorithm employed in this program are provided in Chapter III of this report. WMARQRDT requires that the series to be modeled reside in FILE FT02F001; it should be transformed, if desired, but generally not yet differenced, since the program performs differencing. Inputs at the terminal include number of seasonal and non-seasonal differences, length of the seasonal period, number and type of parameters desired, and initial estimates of those parameters. The program then calculates the least-squares parameter estimates,

the standard error of these estimates, the MA constant term, the residual variance, the autocorrelations and partial autocorrelations of the residuals, the final model sum of squared residuals and a Chi-square goodness-of-fit statistic (the portmanteau test). Additionally, it plots offline the autocorrelations and partial autocorrelations of the residuals. WMARQRDT uses the IMSL subroutines FTRDIF, FTAUTO, LINVLf and MDCDFI, the SSPLIB routine PLOT8, and the Time Series Editor resident subroutines PARSH, MARQRT, SUMSQ, SWAPB and FORMB. The calculations are performed in double-precision arithmetic.

J. XSUMSQ PROGRAM

The XSUMSQ program accepts an already transformed and differenced series in FILE FT02F001 and general seasonal and non-seasonal Box-Jenkins parameter values for a general seasonal model from the terminal. The program then calculates the residual sum of squares for these parameters, giving the user a feel for the "goodness-of-fit" the parameter estimates provide in modeling the given time series. XSUMSQ uses the Time Series Editor resident subprogram XSUMSQ. All output is printed at the terminal. The calculations are performed in double-precision arithmetic.

K. FORECAST PROGRAM

The FORECAST program uses the estimated Box-Jenkins (seasonal or non-seasonal) model to compute forecasts of

the transformed and differenced time series. FORECAST requires that the time series to be forecasted reside in FILE FT02F001 in transformed form. During execution the input required at the terminal includes the forecast origin, the numbers and estimated values of the AR and MA parameters of the model, the overall MA constant, the plot origin index, the maximum forecast lead time, the order of differencing in the model, and a level of significance for the forecast confidence limits. The program needs the transformation parameters from FILE FT07F001, created when using program TRANS, and performs the inverse transform to return the data to its original form. FORECAST also undifferences differenced data. The program output includes the forecasts up to the given maximum lead time, the deviations from each forecast for the confidence limits, and plots of forecasts and confidence limits plotted offline. The FORECAST program uses IMSL subroutine FTCAST and the special IBM SSPLIB subroutine UTPLT8. The plot itself and output data are written onto FILE FT08F001 on the user's disc; since this plot file takes considerable space, it is usually best to erase it after execution of the program is completed and the plot itself has been printed offline.

L. ROOTS PROGRAM

The ROOTS program is used to determine the roots of the characteristic equation for a general ARIMA model. The roots are useful for testing for stationarity and for

determining the form of the forecast function. ROOTS uses the IMSL subroutine ZPOLR to calculate the roots; it requires as input the number of AR parameters in undifferenced form,¹ and the actual values of these parameters, as determined from the ESTIMATE or WMARQRDT programs. The roots are printed out at the terminal. If desired, roots can also be obtained for the MA polynomial, simply by substituting the appropriate MA values for the AR values as input.

M. GENERATE AND SIMULATE PROGRAMS

Because of their similarities, the GENERATE and SIMULATE programs are described together. The GENERATE program permits the user to generate a time series from any non-seasonal ARIMA model he specifies. The user must identify the model and give values for the model parameters and starting conditions; a random number seed must also be input. The program takes the specified model, generates random noise terms, and calculates as many values of the time series as desired. The GENERATE program can be useful for purposes of classroom instruction in the generation of a wide variety of time series examples for model identification exercises.

¹Suppose the model was ARIMA(1,1,0). The differenced form would be $(1-\phi_1 B)\nabla^1 z_t = \theta_0 + a_t$. Then, the undifferenced form would be written $(1-(1+\phi_1)B + \phi_1 B^2)z_t = \theta_0 + a_t$, where the left-hand side was found by multiplying $(1-\phi_1 B)$ by $(1-B) = \nabla^1$.

It could also be useful in the diagnostic phase of model checkout; a time series could be generated from the estimated model, and its properties compared with those of the original series. If large discrepancies occur in this comparison, this may be evidence that the model is inadequate. The model output, the generated time series, is written in FILE FT02F001 as well as being printed offline. The GENERATE program uses IMSL subroutine FTGEN1.

The SIMULATE program provides the capability of generating any number of simulated time series. This program is useful for prediction of what might happen in the future; it also demonstrates that within a given model, the time series actually observed can vary considerably. This program uses GENERATE, but also requires as input the actual series in FILE FT02F001¹ and the number of simulated series the user wishes to generate. The series should be already transformed and differenced. Additionally, SIMULATE allows the user to select values of the original time series as starting values for the simulated series. The output consists of the simulated series printed at the terminal. The SIMULATE program uses IMSL subroutine FTGEN1.

N. HELP OPTION

The HELP option is a modification to the Time Series Editor CP/CMS program itself that allows the user to obtain

¹If desired, the user may choose to input his own starting values, and not use the actual time series values.

information about Editor programs after the initial introductory phase has been completed.

0. SUMMARY

This chapter has provided brief descriptions of the options available for analysis of time series data in the Time Series Editor. The following two chapters contain detailed examples of time series analysis for both a non-seasonal and a seasonal series; the accompanying terminal sessions and computer output for analysis of the seasonal series is contained in Appendix B. Appendix A contains a User's Guide for the Time Series Editor, and Appendix C contains listings of the data sets used in the analysis of the next two chapters.

V. EXAMPLE NON-SEASONAL TIME SERIES ANALYSIS

This chapter provides a description of a complete analysis of a non-seasonal time series using the Time Series Editor. The time series analyzed is Series C (Chemical Process Temperature Readings), taken from Box and Jenkins [Ref. 4, p. 528]. This time series was selected since it is fully analyzed and discussed in Reference 4, so that the interested user can compare for himself the results presented there with the results given by the Time Series Editor. After reading this overview of the analysis, the reader can follow the actual user session and associated user session Editor program output contained in Appendix B to this report.

The first step in the analysis of series C is to plot the series using program PLOT. The plot reveals rather wide fluctuations in the series, but not the sort of explosive nonstationary behavior that would render an attempt to model fruitless. The plot also reveals that the time series has a large amount of momentum (movements of the series tend to resist changes of direction over time). This is characteristic of ARIMA models with two or more differences.

As a second step, the autocorrelations (autos) and partial autocorrelations (pautos), mean, and variance of the series were estimated using AUTO. The plot shows that the autos dampen out slowly in a nearly linear fashion.

This is an indication that the process is nonstationary and that one or more differences are required to make it stationary. The pauto plot is not informative when the autos fail to dampen out rapidly.

As suggested by the plots of the original series and its auto and pautos, the program DIFF was used to examine the series with both one and two non-seasonal differences; i.e., the series $z'_t = \nabla^1 z_t$ and $z''_t = \nabla^2 z_t$ were derived. The plots suggest that two differences might be required to achieve stationarity in the series. However, one might be able to get by with a single difference and an AR parameter near unity, since the autos of z'_t dampen out slowly. Thus, two possible model candidates are suggested:

$$1. \text{ ARIMA}(1,1,0): (1 - \phi_1 B) \nabla^1 z_t = \theta_0 + a_t ,$$

and

$$2. \text{ ARIMA}(0,2,2): \nabla^2 z_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2) a_t .$$

For purposes of estimation, the second model was extended to two moving average parameters. Such overfitting is often done to see if the estimated moving average parameters turn out to be near zero, thus confirming the tentative identification.

The next step in the analysis process is to calculate the maximum likelihood estimates of the model parameters.

The program ESTIMATE was used to do this. The estimated parameters for the ARIMA(1,1,0) model were:

$$\begin{aligned}\hat{\phi}_1 &= 0.8073 \\ \hat{\theta}_0 &= -0.006789 \\ \hat{\sigma}_a^2 &= 0.0177633\end{aligned}$$

The autos and pautos of the residuals were calculated to test the model. The correlations should appear to be estimates of a pure white noise process if the model is adequate; the (1,1,0) model seemed to pass that test. The chi-square goodness-of-fit test results in a $\chi^2 = 28.88$ with 24 df and a significance level of 0.2247. Thus, there is no strong evidence to suggest that the (1,1,0) model is inadequate. The model can be written, in undifferenced form:

$$(1 - 1.8073B + 0.8073B^2)z_t = a_t.$$

The θ_0 is omitted here since it is nearly zero in value.

The estimated parameters for the (0,2,2) model were

$$\begin{aligned}\hat{\theta}_1 &= 0.1378 \\ \hat{\theta}_2 &= 0.1296 \\ \hat{\theta}_0 &= -0.0026787 \\ \hat{\sigma}_a^2 &= 0.0189559\end{aligned}$$

As before, the correlation plots of the residuals fail to suggest any inadequacy of the model. However, the chi-square goodness-of-fit test for this model yielded a $\chi^2 = 36.75$ with 23 df and a significance level of 0.0346. Therefore,

if the ARIMA(0,2,2) model were a "correct" one, a chi-square value as large as 36.75 would occur by chance with a probability of only 0.0346. Due to its simplicity (parsimony is a desirable feature in time series models) and better fit, the ARIMA(1,1,0) model was chosen as the "best" of the two alternatives originally selected.

The ARIMA(1,1,0) model calculated was then used to make forecasts and confidence limits for these forecasts, using program FORECAST. The forecasts were made for 25 periods into the future, with plot origin 200 and forecast origin 220. The length of the series C itself is 226; with one difference, it becomes 225. The plot in Appendix B shows the forecasted values, along with the confidence limits at level of significance 0.10; the plot shows that the width of the confidence limits increases substantially as the lag gets large.

Finally the program ROOTS was used to calculate the roots of the characteristic equation for the ARIMA(1,1,0) model, which can be written:

$$1 - 1.8073B - 0.8073B^2 = 0$$

The roots calculated by ROOTS were 1.239 and 1.0. Also, the program SIMULATE was used to generate a simulated series from the ARIMA(1,1,0) model, using the last two values of the actual series as starting values. The entire session lasted about two hours, including checkout of plots and consumed less than a minute of CPU time.

VI. EXAMPLE SEASONAL TIME SERIES ANALYSIS

This chapter provides a description of a complete analysis of a seasonal time series using the Box-Jenkins technique through the Time Series Editor. The time series analyzed is Series G (International Airline Passengers: Monthly Totals (Thousands of Passengers) January 1949 - December 1960), taken from Box and Jenkins [Ref. 4, p. 531]. This time series was selected since it was fully analyzed and discussed in Chapter 9 of Reference 4; again, the interested reader can compare for himself the results presented by Box and Jenkins with those given by the Time Series Editor programs. After reading this overview of the analysis, the reader can follow the actual user session and associated Editor program output contained in Appendix C to this report.

The plot shows both annual seasonality and an increasing trend. This suggests that a seasonal model with differencing may be needed. The autocorrelations (autos) and partial autocorrelations (pautos) were then calculated and plotted using AUTO. The plot of the autos shows very slow damping out with peaks at intervals of 12, indicating strong non-stationarity and a seasonal period of 12 months. Again, here the pauto plot is not informative, due to the slow damping out of the autos. As it is sometimes useful in dealing with seasonal models, a natural log transform was

made on the series G data using program TRANS. The plot of the logged data shows little change in non-stationarity, and the autos and pautos of the logged data still indicate seasonality and non-stationarity.

Next, differencing to achieve stationarity was tried, using program DIFF; both one seasonal and one non-seasonal difference of the logged data was taken. The autos and pautos of the data $\nabla^1 \nabla_{12}^1 (\ln z_t)$ appear much improved, rapidly damping out. They also exhibit sharp peaks at the periods of seasonality, indicating the need for seasonal parameterization in the model. Since positive identification of the model type from the auto and pauto plots was not possible, two were postulated as close candidates:

1. ARIMA (0,1,1) x (0,1,1)¹², written:

$$\nabla^1 \nabla_{12}^1 (\ln z_t) = \theta_0 + (1-\theta_1 B)(1-\theta_{1,12} B^{12}) a_t,$$

and

2. ARIMA (1,1,1) x (0,1,1)¹², written:

$$(1-\phi_1 B) \nabla^1 \nabla_{12}^1 (\ln z_t) = \theta_0 + (1-\theta_1 B)(1-\theta_{1,12} B^{12}) a_t.$$

The next step in the analysis was to calculate the parameter estimates for both models. YESTSEAS was used to

obtain initial parameter estimates. For the model
ARIMA (0,1,1) x (0,1,1)¹², YESTSEAS yielded starting values:

$$\hat{\theta}_{1(s)} = 0.390425$$

$$\hat{\theta}_{1,12(s)} = 0.534397 .$$

For this model, calculation of final parameter estimates
in WMARQRDT using the starting values above yielded the
following in seven iterations:

$$\hat{\theta}_1 = 0.377152$$

$$\hat{\theta}_{1,12} = 0.572387$$

$$\hat{\theta}_0 = 0.000291 ;$$

the chi-square statistic for residual lack of fit of the
model was $\chi^2 = 29.713571$, with df = 38 and a probability
of exceeding the χ^2 value of .829.

For the ARIMA (1,1,1) x (0,1,1)¹² model YESTSEAS calcu-
lated starting values:

$$\hat{\phi}_{1(s)} = 0.112699$$

$$\hat{\theta}_1(s) = 0.490883$$

$$\hat{\theta}_{1,12(s)} = 0.533679 ;$$

for this model, WMARQRTD yielded the following parameter values in ten iterations:

$$\hat{\phi}_1 = 0.146667$$

$$\hat{\theta}_1 = 0.509816$$

$$\hat{\theta}_{1,12} = 0.573388 .$$

The chi-square statistic for residual lack of fit of the model was $\chi^2 = 28.302543$, with $df = 37$ and a probability of exceeding the χ^2 value of .847.

Next, the plots of autos and pautos of the residual for both models were examined; both models have autos and pautos that dampen out rapidly, and essentially appear to be "white noise", with the exception of a small peak at auto and pauto of lag 12, corresponding to the seasonal period. The chi-square statistics indicate that both models have "about" the same goodness of fit to the data. Consequently, both models were used to make forecasts of the series.

Program FORECAST was used to forecast series G into the future for 30 time periods, with a forecast origin of 131; the plot origin selected was index 100. As expected, both models do about the same job of forecasting the series.

The significance level for the confidence limits for both plots was 90%.

As a final check of both models, the program XSUMSQ was used to calculate the sum of squared residuals for models. The results were as follows:

1. For ARIMA (0,1,1) x (0,1,1)¹² ,

$$S^2(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_{1,12}, \hat{\sigma}_a^2) = 0.181929$$

2. For ARIMA (1,1,1) x (0,1,1)¹² ,

$$S^2(\hat{\phi}_1, \hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_{1,12}, \hat{\sigma}_a^2) = 0.181643 .$$

It can be seen here that both models are nearly equally "good", and either could be used with a reasonable amount of confidence for forecasting series G.

VII. SUMMARY

There has been an increasing need in military applications, as well as in the general business area, for a computer-aided time series analysis and forecasting capability that is both powerful and easy to use. Current military-associated functions in which time series techniques are valuable include material inventory management, recruiting, personnel management, budget analysis, force level projections, and perhaps even short-term weather prediction.

The Time Series Editor that has been described in this report provides a unified collection of programs that greatly facilitate the conduct of a complete time series analysis and forecasting evolution using Box-Jenkins methodology. The routines guide the user from the data input stage, through the data analysis and parameter estimation stages, and finally through the model building, diagnostic checking and forecasting stages. With its simple input requirements, the Editor can be quickly mastered by even a beginning computer user, who has a basic understanding of Box-Jenkins methodology.

The Box-Jenkins methodology has been described broadly in Chapter II and descriptions of the programs resident in the Time Series Editor have been given in Chapter IV. Example analyses of both non-seasonal and seasonal time series are provided in Chapters V and VI, with complete

user sessions and output packages for these analyses included as Appendices B and C. The Appendices also contain a Time Series Editor User's Guide, as well as complete program listings.

Although the Time Series Editor in its present form provides an excellent capability for interactive time series analysis using the Box-Jenkins methodology, there are additions to the program set that might be made in the future which would improve the Editor's capability and utility. It is first recommended that the model diagnostic testing capability be extended to include a periodogram analysis, and/or other tests related to the spectral analysis of time series [see Ref. 4, p. 294]. Another useful addition might be to include in the Time Series Editor a package capable of improving the ease of entering data and simplifying the input error correction problem.

APPENDIX A

USER'S GUIDE TO THE TIME SERIES EDITOR

The Time Series Editor is a collection of FORTRAN programs driven by a control program called TIMESER EXEC in the CP/CMS executive language that has been designed specifically for the analysis and forecasting of time series data using the general Box-Jenkins methodology. This Guide contains information essential for the user to access the Time Series Editor, enter time series data, build a model and obtain the desired output.

I. DATA INPUT

The Time Series Editor requires that time series data be entered into the sequential FORTRAN input/output file named FILE FT02F001. This can be done either online or via cards read offline. The following diagram illustrates the proper card deck arrangement. When your deck is ready for input, give it to a computer center system operator to direct it to your disc space.

If the user has a data deck already punched up in a format other than 5F15.6, he may enter the series as above into FILE FT03F001 (without the length of series card) and use the Editor program ZFORMAT to transform it into the FILE FT02F001 in the proper format, without destroying his original file. This program is described in the next section.

time saver data deck
in FORMAT (5F15.6)

length of series in
FORMAT (I3)
(omit for ZFORMAT
program)

[illegible]

II. TABLE OF OPTIONS

This Table provides the user with the basic information necessary to understand the data requirements, functions and output options for each program in the Editor.

PROGRAM	INPUT	OUTPUT	REMARKS
Name: ZFORMAT Entry Code: z	Files: (1) data in FILE FT03F001 Keyboard: (1) length of time series (2) format of time series	Files: (1) original data put in FORMAT(5F15.6) in FILE FT02F001 (2) original data unchanged in FILE FT03F001	(1) puts date into proper format for Time Series Editor (2) single precision
Name: CMSWORK Entry Code: C	no specific input; normal usage would include file name alteration, obtaining disc status, or erasing files no longer required	no specific output	(1) allows users to perform CMS admin actions while in TIMESER environment
Name: TRANS Entry Code: T	Files: (1) data in FILE FT02F001 Keyboard: (1) origin transformation? (2) scale change factor? (3) log transform? (4) power/root transform?	Files: (1) original data unchanged in DATA FT02F001 (2) transformed data in FILE FT02F001 (3) transformation parameters in FILE FT07F001	(1) allows user to transform data in a file (2) single precision

PROGRAM	INPUT	OUTPUT	REMARKS
Name: DIFF Entry Code: D	Files: (1) original (transformed if desired) series in FILE FT02F001 Keyboard: (1) series seasonal? (2) number of non- seasonal differences (3) number of seasonal differences (4) length of seasonal period	Files: (1) original series unchanged in FILE FT02F001 (2) differenced series in FILE FT03F001	(1) allows user to perform differencing of a time series, in order to achieve series stationarity (2) uses IMSL subroutine FTRDIF (3) single precision
Name: PLOT Entry Code: P	Files: (1) original series in FILE FT02F001 Keyboard: (1) title for plot	Files: (1) original series unchanged in FILE FT02F001 Offline: (1) plot of time series	(1) allows user to plot any time series using offline printer (2) uses SSPLIB routine PLOT8 (3) single precision

PROGRAM	INPUT	OUTPUT	REMARKS
Name: AUTO Entry Code: A	<p>Files:</p> <p>(1) original series in FILE FT02F001 transformed and/or differenced if desired</p> <p>Keyboard:</p> <p>(1) number of autos/pautos to be calculated</p> <p>(2) title for plots of autos/pautos</p>	<p>Files:</p> <p>(1) original series unchanged in FILE FT02F001</p> <p>(2) plots of autos/pautos in FILE FT08F001</p> <p>Terminal:</p> <p>(1) values of autos and pautos</p> <p>(2) mean</p> <p>(3) variance</p> <p>Offline:</p> <p>(1) plots of autos and pautos</p>	<p>(1) allows user to obtain basic statistics for time series</p> <p>(2) uses IMSL subroutine FTAUTO, as well as UTPLT8 routine</p> <p>(3) single precision</p>
Name: ESTIMATE Entry Code: E	<p>Files:</p> <p>(1) original series in FILE FT02F001; transformed and/or differenced as desired.</p> <p>Keyboard:</p> <p>(1) number of AR parameters</p> <p>(2) number of MA parameters</p> <p>(3) number of (non-seasonal) differences to be taken, if series not already differenced</p> <p>(4) titles for plots of residual autos and pautos</p>	<p>Files:</p> <p>(1) original series unchanged in FILE FT02F001</p> <p>(2) model residuals in FILE FT02F001</p> <p>Terminal:</p> <p>(1) estimated AR parameters</p> <p>(2) estimated MA parameters</p> <p>(3) MA constant</p> <p>(4) residual variance</p> <p>(5) portmanteau test of residuals</p> <p>Offline:</p> <p>(1) plots of residual autos and pautos</p>	<p>(1) allows user to obtain maximum likelihood parameter estimates for a general non-seasonal ARIMA model, as well as data concerning model sufficiency</p> <p>(2) uses IMSL subroutine FTMAXL</p> <p>(3) single precision</p>

PROGRAM	INPUT	OUTPUT	REMARKS
Name: YESTSEAS Entry Code: Y	Files: (1) original series in FILE FT02F001 transformed if desired; and either differ- enced or undifferenced Keyboard: (1) number of seasonal and non-seasonal differences to be taken (2) length of seasonal period (3) numbers of both seasonal and non- seasonal AR and MA parameters	Files: (1) original series unchanged in FILE FT02F001 Terminal: (1) estimated values for requested seasonal and non-seasonal AR and MA parameters	(1) allows user to calcu- late initial non- seasonal and seasonal ARIMA model parameter estimates as input to WVARQORDT routine (2) uses IMSL subroutines FTRDIF, FTMAXI, FTAUTO, FTARPS and FTMAPS (3) single precision
Name: WVARQORDT Entry Code: W	Files: (1) original series in FILE FT02F001; may be transformed but not differen- ced (either diff or nondiff) Keyboard: (1) number of seasonal and non-seasonal differences to be taken (2) length of seasonal period	Files: (1) original series unchanged in FILE FT02F001 Terminal: (1) seasonal and non- seasonal AR and MA parameter estimates parameter standard errors (3) MA constant term (4) sum of squared residuals (5) residual variance	(1) allows user to estimate non-linear least squares parameters for a general seasonal Box-Jenkins ARIMA model (2) requires initial parameter estimates, obtainable using YESTSEAS program (3) uses IMSL subroutines FTRDIF, FTAUTO, LINVLF, and MDXDFI, SSPLIB routine DPLOT, and Time Series Editor resident subroutine PARSH, MAROPT, SUMSQ, SWAPB and FORMB

PROGRAM	INPUT	OUTPUT	REMARKS
Name: WVARQORDT (CONTINUED)	(3) number and initial estimates of seasonal and non-seasonal AR and MA parameters	(6) residual variance (7) portmanteau test for model goodness-of-fit Offline: (1) plots of autops of residuals	(4) double precision (5) requires LOGIN with 450k core (6) if non-seasonal modeling, input length of season = 1
Name: XSUMSQ Entry Code: X	Files: (1) original series in FILE FT02F001; should be transformed and/or differenced as desired Keyboard: (1) number and estimates of seasonal and non-seasonal AR and MA parameters (2) length of seasonal period	Files: (1) original series unchanged in FILE FT02F001 Terminal: (1) value of residual sum of squares for the model and parameters specified	(1) allows user to obtain a residual sum of squares value for any seasonal or non-seasonal ARIMA model with specified parameters (2) uses Time Series Editor resident subroutine SUMSQ (3) double precision
Name: FORECAST Entry Code: F	Files: (1) original series in FILE FT02F001; transformed as desired, but not differenced Keyboard: (1) number of seasonal and non-seasonal differences to be taken	Files: (1) original series unchanged in FILE FT02F001 (2) forecast values and plot FILE FT08F001	(1) allows the user to forecast any seasonal or non-seasonal time series using a previously determined seasonal or non-seasonal ARIMA model and the time series itself

PROGRAM	INPUT	OUTPUT	REMARKS
Name: FORECAST (CONTINUED)	<p>(2) length of seasonal period</p> <p>(3) number and estimated values of seasonal and non-seasonal AR and MA parameters</p> <p>(4) MA constant</p> <p>(5) index for forecast origin</p> <p>(6) maximum forecast lead time</p> <p>(7) index for plot origin</p> <p>(8) confidence level for forecast confidence units</p>	<p>Offline:</p> <p>(1) plot of forecast of series, including listing of forecast values and confidence interval values</p>	<p>(2) uses IMSL subroutine FTRDLF and modified SSPLTB routine</p> <p>UTPLOT called</p> <p>UTPLT8</p> <p>(3) single precision</p>
Name: ROOTS Entry Code: R	<p>Keyboard:</p> <p>(1) number of AR parameters in undifferenced form</p> <p>(2) values of AR parameters</p> <p>None</p>	<p>Terminal:</p> <p>(1) values of roots</p> <p>None</p>	<p>(1) allows user to calculate the roots of the characteristic equation for non-seasonal ARIMA models</p> <p>(2) uses IMSL routine ZPOLR</p> <p>(3) single precision</p>
Name: HELP Entry Code: H	None	None	<p>(1) allows user to access program information paragraphs after the TIMESER introduction phase has been completed</p>

PROGRAM	INPUT	OUTPUT	REMARKS
Name: GENERATE Entry Code: G	<p>Keyboard:</p> <ol style="list-style-type: none"> (1) random number seed (2) number and values for non-seasonal AR and MA ARIMA model parameters (3) MA constant term (4) residual variance (5) length of series to be generated (6) initial starting value for time series to be generated 	<p>Files:</p> <ol style="list-style-type: none"> (1) generated time series written onto FILE FT02F001 <p>Offline:</p> <ol style="list-style-type: none"> (1) length of generated series and series values themselves are printed offline 	<ol style="list-style-type: none"> (1) allows the user to generate a time series from a given non-seasonal ARIMA model, previously determined (2) uses IMSL subroutine FTGEN1 (3) single precision
Name: SIMULATE Entry Code: S	<p>Files:</p> <ol style="list-style-type: none"> (1) original series in FILE FT02F001, transformed and/or differenced as desired <p>Keyboard:</p> <ol style="list-style-type: none"> (1) number and values for non-seasonal AR and MA ARIMA model parameters (2) MA constant term (3) residual variance (4) index value of time series where simulation is to begin (5) starting values of series to be simulated (6) random number seed (7) number of values to be simulated in each series (8) number of series to 	<p>Files:</p> <ol style="list-style-type: none"> (1) original series unchanged in FILE FT02F001 <p>Terminal:</p> <ol style="list-style-type: none"> (1) Simulated series 	<ol style="list-style-type: none"> (1) allows user to produce any number of simulated time series from a given non-seasonal ARIMA model (2) uses IMSL subroutine FTGEN1 (3) single precision

III. THE BASIC USER SESSION

To use the Time Series Editor, the user must log into CMS, get into CP, link to the disc storage area where the Time Series Editor resides, reimplement CMS, log into the general user and Time Series Editor disc areas, and enter the TIMESER routine. This section will provide explicit guidelines to enable the user to perform the above steps on the NPS CP/CMS system. Commands marked with an asterisk (*) are those actually entered on the terminal by the user (the asterisk itself is omitted). Those without an asterisk and those written in all capital letters are system responses at the terminal. Numbered sentences are comments, which will not appear during an actual user session. The instructions and system responses assure the user is on an IBM 2741 Input/Output Terminal. Some minor modifications may be necessary if other terminals are used.

1. Turn the terminal on, depress the BREAK key, and wait for the system to respond:

```
CP-67 online xd.65 gsyosu
```

2. Depress the ATTN key. The roll bar will advance and the keyboard will unlock. Then enter:

```
*login aaaapbb 450k
```

3. aaaa is the user's identification number, and nn is the terminal number (usually written on the terminal). For example, if the user's ID number is 1621 and the terminal number is 44, the input would be:
login 1621p44 450k. The addition of 450 k to the normal login command is necessary to execute the program WMARQRDT in the Editor; for users not planning to execute this program during a session, this addition is not necessary.

4. The system will respond with the statement:
ENTER PASSWORD:
5. The user then enters his password, or the general users password npg;
*password
6. The system will then respond:
ENTER 4-DIGIT PROJECT NUMBER FOLLOWED BY 4-CHARACTER
COST CENTER CODE:
7. The user then enters:
*gggghhhh
8. gggg is the assigned project number, and hhhh is the user's section designator or the faculty code.
9. The system will respond with the message of the day, such as:

CP/CMS HOURS ... 0930=2200(MON-THURS) ... 0930-1800(FRI)
OUTPUT RETAINED 5 DAYS
Cms Version 3.25
10. At this point, the user is in CMS. He must then get into CP; this can be done by hitting the ATTN key. The system will then respond:

CP
11. The user must then link to the TIME SERIES EDITOR; this is accomplished by entering:

*link 2069p 191 192
12. The system will respond with:

ENTER PASSWORD:
13. The password (read only) to enter the Editor is:

*timser
14. The system then responds:

SET TO READ ONLY

15. The user now implements CMS by:
- *ipl cms
16. The system will respond:
- CMS Version 3.25
17. Now the user must log into both the general user and the Time Series Editor area by entering:
- *login 191
18. The system will respond with a message such as:
- R;
19. The user then enters the command:
- *login 192 t,p
20. The system will respond:
- T (192) R/O
R;
21. The user can then enter the Time Series Editor (guided version) by entering the command:
- *timeser
22. The system will respond:
- EACH 2 SECONDS EXECUTION TIME IS INDICATED BY *
- YOU HAVE ENTERED THE TIME SERIES EDITOR
- PLEASE RESPOND TO EACH QUERY WITH AN INPUT AT THE TERMINAL.
ENTER ONLY THE FIRST LETTER FOR A WORD RESPONSE.
ENTER NUMERICAL VALUES VIA FORTRAN FORMAT..
- TYPE INTEGER VALUES (RIGHT JUSTIFIED) FOR NAMES STARTING
WITH I THROUGH N. TYPE FLOATING VALUES WITH DECIMAL FOR
ALL OTHERS.
- DO YOU WANT A LIST OF THE OPTIONS?
23. The user is then on his own, guided by the Exec routine. See the notes that appear at the end of this guide for additional information. Eventually the user will be asked:
- DO YOU WANT TO TRY AGAIN?

24. If a yes response is given, another sequence will begin; if the response is no, the user will be taken out of the Time Series Editor environment and returned to CMS. The system response will be:

CONTROL RETURNED TO CMS
R;

25. The user can then log out of CMS by typing:

*cp logout

26. The system will respond with:

CONNECT= 00:08:02 VIRTCPU= 000:07.98 TOTCPU= 000.10.94
LOGOUT AT 14.22.04 on 10/16/78

27. The user should then turn off his terminal and tear off the output from his session.

The more experienced user can dispense with the "welcome aboard" section of the Time Series Editor and get right down to business by using the shortened version of the Editor. This shortened version may be entered by linking in the normal way, and then entering the Editor by typing the COMMAND

*timeser s (asterisk omitted).

The system will immediately respond:

ENTER LETTER FOR OPTION YOU WANT.

The session inside the Editor then begins.

IV. BRIEF SAMPLE USER SESSION

A brief sample user session is given below; it includes copies of the offline output generated during the session.

repeat login nur@pn@

login 1621p44 450k
ENTER PASSWORD:

ENTER 4-DIGIT, PROJECT NUMBER FOLLOWED BY 4-CHARACTER COST CENTER CODE:
0444r172
READY AT 17.24.38 ON 09/16/78
CMS Version 3.25

stat
P (191): 29 FILES; 241 REC IN USE, 55 LEFT (of 296), 83% FULL (2 CYL)
R;
cp q f
FILES:- NO RDR, NO PRT, NO PUN
R;
CP
link 2069p 191 192
ENTER PASSWORD:

SET TO READ ONLY

ipl cms
CMS Version 3.25

login 191
R;
login 192 t,p
T (192) R/O
R;

timeser
EACH 2 SECONDS EXECUTION TIME IS INDICATED BY *.

YOU HAVE ENTERED THE TIME SERIES EDITOR.

PLEASE RESPOND TO EACH QUERY WITH AN INPUT AT THE TERMINAL.
ENTER ONLY THE FIRST LETTER FOR A WORD RESPONSE.
ENTER NUMERICAL VALUES VIA FORTRAN FORMAT.

TYPE INTEGER VALUES (RIGHT JUSTIFIED) FOR NAMES STARTING
WITH I THRU N. TYPE FLOATING VALUES WITH DECIMAL FOR ALL OTHERS.

DO YOU WANT A LIST OF THE OPTIONS?
Y

OPTION	DESCRIPTION
GENERATE	-----GENERATE ANY ARIMA TIME SERIES
AUTO	-----CALCULATE AUTOCORRELATIONS, PAUTOS, MEAN AND VARIANCE
PLOT	-----PLOT A TIME SERIES
ESTIMATE	-----CALCULATE MAX LIKELIHOOD ESTIMATES OF ARMA PARAMETERS
DIFF	-----DIFFERENCE A TIME SERIES
FORECAST	-----FORECAST FUTURE VALUES, CONSTRUCT CONFIDENCE INTERVALS
TRANS	-----TRANSFORMS VALUES OF A TIME SERIES
ROOTS	-----DETERMINES ROOTS OF ARIMA CHARACTERISTIC EQUATION
ZFORMAT	-----ALTER DATA FILE TO FORMAT 5F15.6
CMSWORK	-----PERFORM CP/CMS COMMANDS IN TIMESER EXEC
SIMULATE	-----SIMULATE NONSEASONAL TIME SERIES
YESTSEAS	-----CALCULATE INITIAL SEASONAL PARAMETERS
MARQRDT	-----MARQUARDT SOLUTION FOR PARAMETER ESTIMATES
XSUMSQ	-----CALCULATE SUM OF SQUARES FOR ARBITRARY PARAMETERS

WOULD YOU LIKE MORE INFO?
Y

ENTER OPTION YOU WANT INFO ABOUT.

a

AUTO -----THIS PROGRAM CALCULATES AUTOCORRELATIONS, PARTIAL
AUTOCORRELATIONS, THE MEAN AND THE VARIANCE FOR A GIVEN TIME SERIES
WHICH MUST RESIDE IN FILE FT02F001. THE PROGRAM USES
FTAUTO IN THE IMSL LIBRARY. THE AUTOCORRELATIONS AND PAUTOS CAN BE
PLOTTED OFFLINE.

DO YOU WANT INFO ABOUT ANOTHER OPTION?

n

DO YOU WANT TO TRY A SESSION?

y

ENTER LETTER FOR OPTION YOU WANT.

c

ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

listf * ft02f001

FILENAME FILETYPE MODE NO.REC. DATE

SERG FT02F001 P1 3 9/16

SERC FT02F001 P1 5 9/16

LNSERG FT02F001 P1 3 9/16

FILE FT02F001 P1 1 9/16

erase file ft02f001

alter serc ft02f001 pl file ft02f001 pl

stat

P (191): 28 FILES; 240 REC IN USE, 56 LEFT (of 296), 81% FULL (2 CYL)

&goto -ques

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

a

IS YOUR DATA IN FILE FT02F001?

y

EXECUTION BEGINS...

AUTOCORRELATIONS

0.978	0.944	0.902	0.854	0.802	0.748	0.692	0.635	0.579	0.923
0.468	0.413	0.359	0.305	0.253	0.201	0.150	0.098	0.047	-0.003
-0.052	-0.101	-0.151	-0.200	-0.248					

PARTIAL AUTOCORRELATIONS

0.978	-0.260	-0.157	-0.093	-0.058	-0.045	-0.012	-0.038	-0.022	-0.010
-0.036	-0.041	-0.038	-0.024	-0.037	-0.027	-0.032	-0.070	-0.048	-0.024
-0.034	-0.061	-0.079	-0.048	-0.037					

MEAN= 22.9739

VARIANCE = 4.22273

ENTER TITLE FOR PLOTS.

autos and pautos for series c data

YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.

PICK UP IN ROOM 1140 UNDER YOUR USER ID NUMBER.

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

c

ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

offline print file ft02f001

alter file ft02f001 pl serc ft02f001 pl

&goto -ques

DO YOU WANT TO GO AGAIN?

n

CONTROL RETURNED TO CMS

R;

AUTOCORRELATIONS

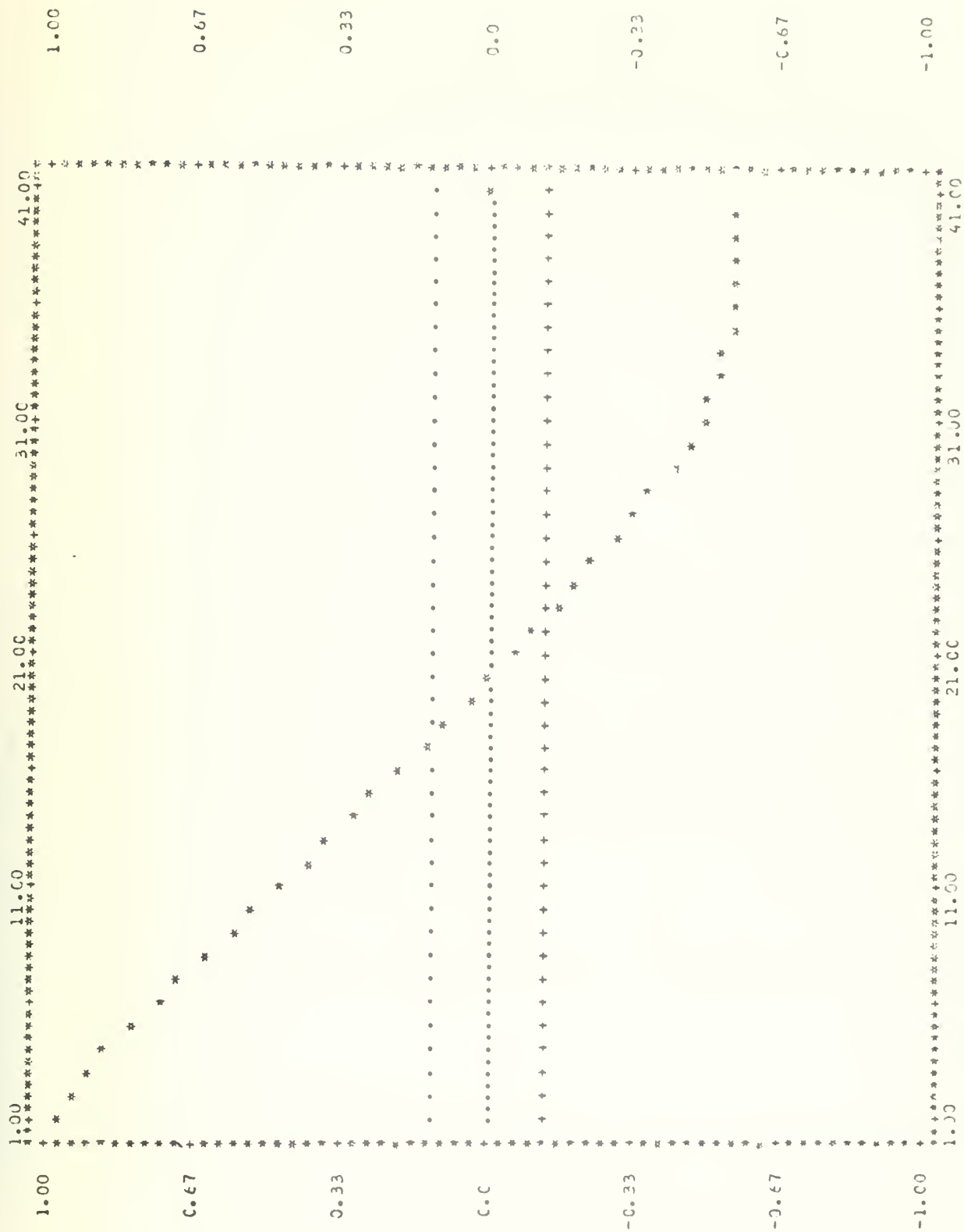
C.578	C.544	0.502	C.854	C.902	C.748	0.692	0.635	0.573	0.523
0.468	C.413	0.353	0.305	0.253	C.201	C.150	0.058	0.047	-0.003
-C.052	-C.101	-0.151	-0.200	-0.248	-0.294	-0.337	-0.379	-0.418	-0.454
-0.486	-C.512	-0.534	-0.550	-0.562	-0.570	-0.573	-0.573	-0.568	-0.562

PARTIAL AUTOCORRELATIONS

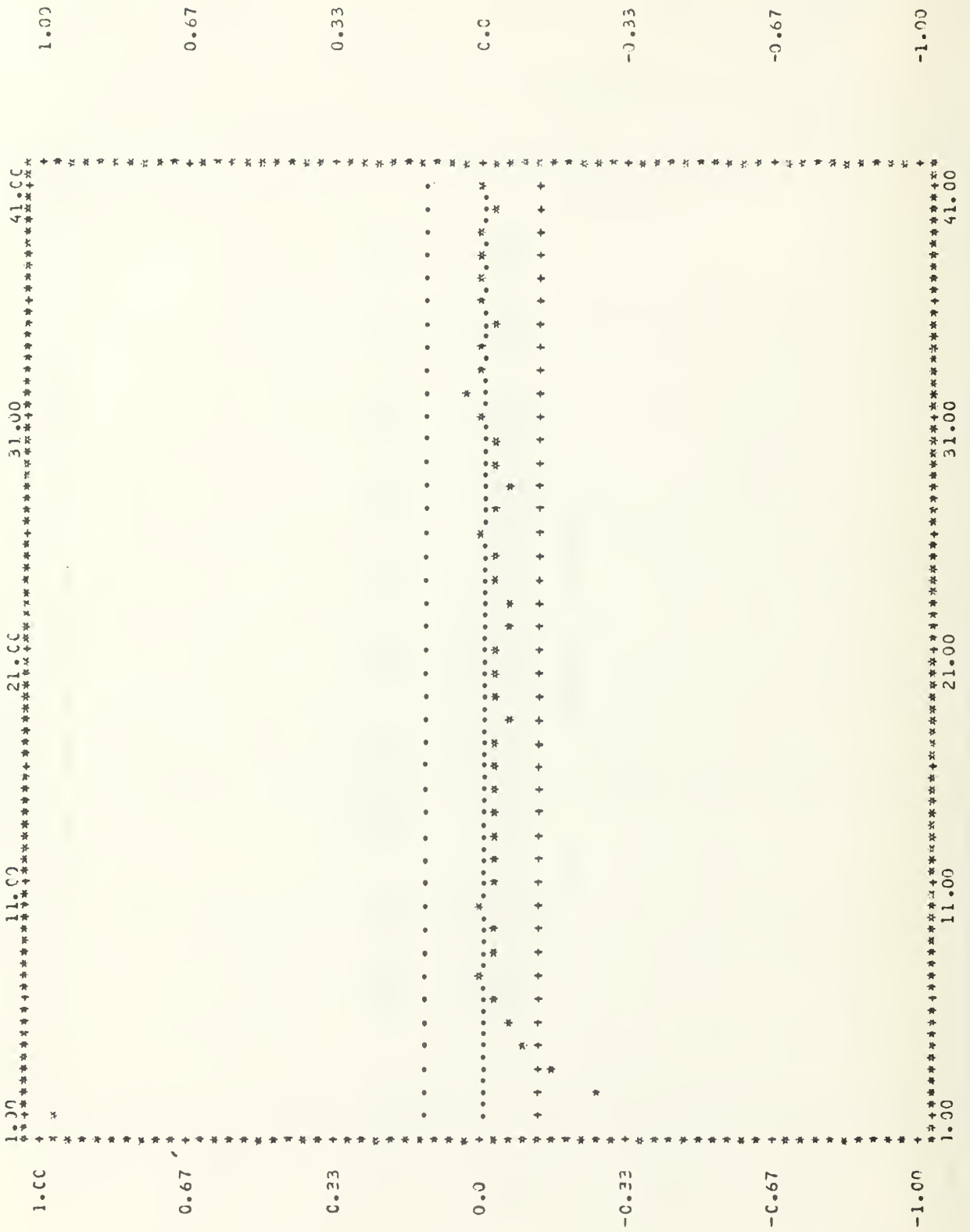
C.578	-0.260	-0.157	-0.093	-0.058	-0.045	-0.012	-0.038	-0.022	-0.010
-0.036	-0.041	-0.038	-0.037	-0.027	-0.027	-0.032	-0.070	-0.048	-0.024
-0.034	-0.061	-0.079	-0.046	-0.037	-0.011	-0.031	-0.054	-0.036	-0.024
-0.002	0.022	-0.014	-0.006	-0.024	0.006	-0.016	-0.004	-0.010	-0.036

MEAN = 22.9739

VARIANCE = 4.22273



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCCORRELATIONS WITH 2 SIGMA BANDS.
AUTOS AND PAUTOS FOR SERIES C DATA

V. PROBLEM CONTROL NOTES

This section will cover corrective measures that can be taken when things fail to go as expected while in the TIMESER environment.

a. A typing error in the CMS environment can be corrected by typing the @ character as many times as is required to back up and then type the correct values. For example, if the user typed timesre, the user could correct the mistake by typing two @ signs, followed by the correct spelling er, as follows: timesre@@re. An entire line can be deleted by typing the ¢ character (or [on some terminals).

b. When working with TIMESER executive programs, the user should exercise care before hitting the return key (or control s on some terminals). If an input value is required and the return key is hit before the proper response is entered, the user will be likely to get thrown out of the editor and have to begin again. In most cases, errors can be corrected only before the return key is struck (in some programs you get a second chance for input).

c. Particular care should be taken for integer value input, which must be right justified in the format field. The editor will advise the user in all cases where the integer format is other than I1.

d. If for any reason the user finds himself in a debug or error condition (caused by erroneous data, or a "blowup" in one of the non-linear optimization routines usually

caused by very poor initial input values), the following procedure will get the user back into the normal TIMESER environment:

- (1) depress the ATTN key twice; this gets the user into CP; hit the ATTN key again, and then type kx to kill the execution.
- (2) re-ipl CMS, and login 191 and then login 192 t,p.
- (3) then the user can type TIMESER or TIMESER S, and return to the TIMESER environment

On the next page is a sample user session where an error causing a debug condition has occurred.

Here the user executed a program that required data in FILE FT02F001, and the FILE did not exist. As the example shows, recovery is quick. Simply hit the break button to get into CP, login 191, login 192 t,p, and then type timeser s to return immediately to the Editor environment.

ENTER LETTER FOR OPTION YOU WANT.

d

EXECUTION BEGINS....

IHC218I FIOCS - I/O ERROR BSAM INPUT ERROR 01 ON FILE: "FT02F001"

8

TRACEBACK ROUTINE	CALLED FROM	ISN	REG.	14	REG.	15	REG.	0	REG.	1
IBCOM			000131FC		000133F8		FFF93908		00014F80	

DIFF CP

ipl cms

CMS Version 3.25

login 191

R;

login 192 t,p

T (192) R/O

R;

timeser s

ENTER LETTER FOR OPTION YOU WANT.

c

ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

alter serc ft02f001 pl file ft02f001 pl

&goto -ques

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

d

APPENDIX 0

KEYBOARD SESSION AND OUTPUT FROM THE ANALYSIS OF TIME SERIES G (SEASONAL)

This appendix contains the computer terminal keyboard session and all offline output from the analysis of Box and Jenkins' [Ref. 4] time series G (International airline passengers, monthly totals, Jan 1949 - Dec 1960) using the Time Series Editor. This appendix is intended to supplement the discussion of the analysis of series G given in Chapter VI of this report.

The terminal listings are presented first, followed by the computer offline output.

repeat login nur@pn@

login 1621p44 400k
ENTER PASSWORD:

ENTER 4-DIGIT PROJECT NUMBER FOLLOWED BY 4-CHARACTER COST CENTER CODE:
0444r172
READY AT 15.12.34 ON 09/16/78
CMS Version 3.25

CP
link 2069p 191 192
ENTER PASSWORD:

SET TO READ ONLY

ipl cms
CMS Version 3.25

login 191
R;
login 192 t,p
T (192) R/O
R;

timeser
EACH 2 SECONDS EXECUTION TIME IS INDICATED BY *.

YOU HAVE ENTERED THE TIME SERIES EDITOR.

PLEASE RESPOND TO EACH QUERY WITH AN INPUT AT THE TERMINAL.
ENTER ONLY THE FIRST LETTER FOR A WORD RESPONSE.
ENTER NUMERICAL VALUES VIA FORTRAN FORMAT.

TYPE INTEGER VALUES (RIGHT JUSTIFIED) FOR NAMES STARTING
WITH I THRU N. TYPE FLOATING VALUES WITH DECIMAL FOR ALL OTHERS.

DO YOU WANT A LIST OF THE OPTIONS?
Y

OPTION	DESCRIPTION
GENERATE	-----GENERATE ANY ARIMA TIME SERIES
AUTO	-----CALCULATE AUTOCORRELATIONS, PAUTOS, MEAN AND VARIANCE
PLOT	-----PLOT A TIME SERIES
ESTIMATE	-----CALCULATE MAX LIKELIHOOD ESTIMATES OF ARMA PARAMETERS
DIFF	-----DIFFERENCE A TIME SERIES
FORECAST	-----FORECAST FUTURE VALUES, CONSTRUCT CONFIDENCE INTERVALS
TRANS	-----TRANSFORMS VALUES OF A TIME SERIES
ROOTS	-----DETERMINES ROOTS OF ARIMA CHARACTERISTIC EQUATION
ZFORMAT	-----ALTER DATA FILE TO FORMAT 5F15.6
CMSWORK	-----PERFORM CP/CMS COMMANDS IN TIMESER EXEC
SIMULATE	-----SIMULATE NONSEASONAL TIME SERIES
YESTSEAS	-----CALCULATE INITIAL SEASONAL PARAMETERS
WMARQDRT	-----MARQUARDT SOLUTION FOR PARAMETER ESTIMATES
XSUMSQ	-----CALCULATE SUM OF SQUARES FOR ARBITRARY PARAMETERS

WOULD YOU LIKE MORE INFO?
Y

ENTER OPTION YOU WANT INFO ABOUT.
Y

YESTSEAS ----THIS PROGRAM CALCULATES INITIAL PARAMETER ESTIMATES
FOR SEASONAL AND NONSEASONAL ARIMA MODELS, TO BE USED AS INPUTS TO
THE WMARQDRT PROGRAM.

DO YOU WANT INFO ABOUT ANOTHER OPTION?
Y
ENTER OPTION YOU WANT INFO ABOUT.
W

WMARQRTD ----THIS PROGRAM CALCULATES NON-LINEAR LEAST SQUARES ESTIMATES OF BOX-JENKINS PARAMETERS FOR A GENERAL (SEASONAL OR NONSEASONAL) ARIMA MODEL. IT REQUIRES INITIAL PARAMETER ESTIMATES AS STARTING VALUES; THESE MAY BE CALCULATED USING PROGRAM YESTSEAS.

DO YOU WANT INFO ABOUT ANOTHER OPTION?

y

ENTER OPTION YOU WANT INFO ABOUT.

x

XSUMSQ -----THIS PROGRAM CALCULATES THE SUM OF SQUARED RESIDUALS FOR ANY SET OF ARIMA (SEASONAL OR NONSEASONAL) PARAMETERS.

DO YOU WANT INFO ABOUT ANOTHER OPTION?

n

*DO YOU WANT TO TRY A SESSION?

y

ENTER LETTER FOR OPTION YOU WANT.

c

ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

stat

P (191): 26 FILES; 235 REC IN USE, 61 LEFT (of 296), 79% FULL (2 CYL)

cp q f

FILES:- NO RDR, NO PRT, NO PUN

listf * ft02f001

FILENAME FILETYPE MODE NO.REC. DATE

SERG FT02F001 pl 3 9/16

SERC FT02F001 pl 5 9/16

alter serg ft02f001 pl file ft02f001 pl

&goto -ques

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

p

IS YOUR DATA IN FILE FT02F001?

y

EXECUTION BEGINS...

ENTER TITLE FOR PLOT

series g data / undifferenced

TIME SERIES PLOTS HAVE BEEN PRINTED OFFLINE

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

a

IS YOUR DATA IN FILE FT02F001?

y

EXECUTION BEGINS...

AUTOCORRELATIONS

0.948	0.876	0.807	0.753	0.714	0.682	0.663	0.656	0.671	0.703
0.743	0.760	0.713	0.646	0.586	0.538	0.500	0.469	0.450	0.442
0.457	0.482	0.517	0.532	0.494					

PARTIAL AUTOCORRELATIONS

0.948	-0.229	0.038	0.094	0.074	0.008	0.126	0.090	0.232	0.166
0.171	-0.135	-0.540	-0.027	0.091	0.025	0.032	0.077	0.048	-0.046
0.046	-0.100	0.052	0.048	-0.163					

MEAN= 280.299

VARIANCE = 14292.0

ENTER TITLE FOR PLOTS.

autos and pautos of series g data / undifferenced
YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.

PICK UP IN ROOM 1140 UNDER YOUR USER ID NUMBER.

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

t

IS YOUR DATA IN FILE FT02F001?

y

EXECUTION BEGINS...

DO YOU WANT TO TRANSLATE THE ORIGIN: W=Z-SHIFT?

n

DO YOU WANT TO RESCALE THE VALUES: W=Z*SCALE?

n

DO YOU WANT A LOG TRANSFORMATION?

y

TRANSFORMATION IS $W(I) = \log(\text{SCALE} * (Z(I) - \text{SHIFT}) - \text{FACTOR})$ WHERE:

SCALE= 1.00000

SHIFT=0.0

FACTOR=0.0

YOUR ORIGINAL SERIES IS IN DATA FT02F001 Pl.

YOUR TRANSFORMED SERIES IS IN FILE FT02F001 Pl.

DO YOU WANT TO PLOT THE TRANSFORMED VALUES?

y

EXECUTION BEGINS...

* ENTER TITLE FOR PLOT

plot of series g data / natural log transform

TIME SERIES PLOTS HAVE BEEN PRINTED OFFLINE

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

a

IS YOUR DATA IN FILE FT02F001?

y

EXECUTION BEGINS...

AUTOCORRELATIONS

0.954	0.899	0.851	0.808	0.779	0.756	0.738	0.727	0.734	0.744
0.758	0.762	0.717	0.663	0.618	0.576	0.544	0.519	0.501	0.490
0.498	0.506	0.517	0.520	0.484					

PARTIAL AUTOCORRELATIONS

0.954	-0.118	0.054	0.024	0.116	0.044	0.038	0.100	0.204	0.064
0.106	-0.042	-0.485	-0.034	0.042	-0.044	0.028	0.037	0.042	0.014
0.073	-0.033	0.061	0.031	-0.194					

MEAN= 5.54218

VARIANCE = 0.193531

ENTER TITLE FOR PLOTS.

autos and pautos of series g data / natural log transform
YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.

PICK UP IN ROOM 1140 UNDER YOUR USER ID NUMBER.

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

d
IS YOUR DATA IN FILE FT02F001?

y
EXECUTION BEGINS...
IS YOUR TIME SERIES SEASONAL?

y
ENTER ORDER OF SEASONAL DIFFERENCING.

1
ENTER LENGTH OF SEASONAL PERIOD VIA I2.

12
ENTER NUMBER OF NONSEASONAL DIFFERENCES.

1
DO YOU WANT TO PLOT AUTO AND PAUTO OF TRANSFORMED DATA?

y
EXECUTION BEGINS...

AUTOCORRELATIONS

-0.341	0.105	-0.202	0.021	0.056	0.031	-0.056	-0.001	0.176	-0.076
0.064	-0.387	0.152	-0.058	0.150	-0.139	0.071	0.016	-0.011	-0.117
0.039	-0.091	0.223	-0.018	-0.100					

PARTIAL AUTOCORRELATIONS

-0.341	-0.013	-0.193	-0.125	0.033	0.035	-0.060	-0.020	0.226	0.043
0.047	-0.339	-0.109	-0.077	-0.022	-0.140	0.026	0.115	-0.013	-0.167
0.132	-0.072	0.143	-0.067	-0.103					

MEAN=0.290920E-03 VARIANCE = 0.208604E-02

ENTER TITLE FOR PLOTS.

autos and pautos for series g s@data / 1 ns diff, 1 seas diff, 1n xform
YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.
PICK UP IN ROOM I140 UNDER YOUR USER ID NUMBER.

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

c
ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.

WHEN FINISHED TYPE: &GOTO -QUES

alter file ft03f001 pl save ft03f001 pl

alter data ft02f001 pl serg ft02f001 pl

stat

P (191): 29 FILES; 267 REC IN USE, 29 LEFT (of 296), 90% FULL (2 CYL)

erase file ft08f001

cp q f

FILES:- NO RDR, NO PRT, NO PUN

&goto -ques

DO YOU WANT TO GO AGAIN?

y

ENTER LETTER FOR OPTION YOU WANT.

y
IS YOUR DATA IN FILE FT02F001?

y
EXECUTION BEGINS...

ENTER LENGTH OF SEASON VIA FORMAT I2.

12
ENTER NUMBER OF NON-SEASONAL DIFFERENCES.

1
ENTER NUMBER OF SEASONAL DIFFERENCES.

1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.

0

ENTER NUMBER OF SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
1
ENTER NUMBER OF SEASONAL MA PARAMETERS.
1

INITIAL PARAMETER ESTIMATES FOR MARQRT :

THETA(1) = 0.390425
THETAS(1) = 0.534397

DO YOU WANT TO GO AGAIN?
Y

ENTER LETTER FOR OPTION YOU WANT.

W
IS YOUR DATA IN FILE FT02F001?

Y
EXECUTION BEGINS...
ENTER NUMBER OF NON-SEASONAL DIFFERENCES.

1
ENTER NUMBER OF SEASONAL DIFFERENCES.

1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.

0
ENTER NUMBER OF SEASONAL AR PARAMETERS.

0
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.

1
ENTER NUMBER OF SEASONAL MA PARAMETERS.

1
ENTER LENGTH OF SEASON VIA FORMAT I2.

12

NOW INPUT YOUR INITIAL PARAMETER ESTIMATES, AS REQUESTED.

ENTER NON-SEASONAL MA PARAMETER THETA(1).
.390425

ENTER SEASONAL MA PARAMETER THETAS(1).
.534397

*

CONVERGENCE HAS BEEN REACHED IN MAX(7, 7) ITERATIONS.

SELECTED OUTPUT FOLLOWS:

PARAMETER ESTIMATE	STANDARD ERROR
THETA(1) = 0.377152	0.819933D-01
THETAS(1) = 0.572387	0.780252D-01

MOVING AVERAGE CONSTANT: THETA0 = 0.000291

**

CHI-SQUARE STATISTIC FOR RESIDUAL LACK OF FIT = 29.713571

DEGREES OF FREEDOM = 38

PROBABILITY OF EXCEEDING STATISTIC = 0.829438

DO YOU WANT TO PLOT AUTO AND PAUTO OF RESIDUALS?

Y

EXECUTION BEGINS...

AUTOCORRELATIONS

0.009	0.026	-0.129	-0.105	0.078	0.077	-0.036	-0.033	0.103	-0.050
0.026	-0.020	0.013	0.035	0.067	-0.130	0.052	0.015	-0.093	-0.090
-0.027	-0.014	0.213	0.009	-0.042					

PARTIAL AUTOCORRELATIONS

0.009	0.026	-0.130	-0.104	0.088	0.068	-0.072	-0.029	0.151	-0.057
-0.022	0.018	0.046	0.001	0.053	-0.114	0.069	0.029	-0.129	-0.134
0.042	-0.014	0.143	-0.025	0.013					

MEAN=0.236005E-02

VARIANCE = 0.139713E-02

ENTER TITLE FOR PLOTS.

autos and pautos of residuals / (0,1,1)x(0,1,1)12 model of series g

YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.

PICK UP IN ROOM 1140 UNDER YOUR USER ID NUMBER.

*

DO YOU WANT TO GO AGAIN?

Y

ENTER LETTER FOR OPTION YOU WANT.

Y

IS YOUR DATA IN FILE FT02F001?

Y

EXECUTION BEGINS...

ENTER LENGTH OF SEASON VIA FORMAT I2.

12

ENTER NUMBER OF NON-SEASONAL DIFFERENCES.

1

ENTER NUMBER OF SEASONAL DIFFERENCES.

1

ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.

1

ENTER NUMBER OF SEASONAL AR PARAMETERS.

0

ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.

1

ENTER NUMBER OF SEASONAL MA PARAMETERS.

1

INITIAL PARAMETER ESTIMATES FOR MARQRT :

PHI(1) = 0.112699

THETA(1) = 0.490883

THETAS(1) = 0.533679

DO YOU WANT TO GO AGAIN?

Y

ENTER LETTER FOR OPTION YOU WANT.

W

IS YOUR DATA IN FILE FT02F001?

Y

EXECUTION BEGINS...

ENTER NUMBER OF NON-SEASONAL DIFFERENCES.

1

ENTER NUMBER OF SEASONAL DIFFERENCES.

1

ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.

1

ENTER NUMBER OF SEASONAL AR PARAMETERS.
 0
 ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
 1
 ENTER NUMBER OF SEASONAL MA PARAMETERS.
 1
 ENTER LENGTH OF SEASON VIA FORMAT I2.
 12
 NOW INPUT YOUR INITIAL PARAMETER ESTIMATES, AS REQUESTED.
 ENTER NON-SEASONAL AR PARAMETER PHI(1).
 0.112699
 ENTER NON-SEASONAL MA PARAMETER THETA(1).
 0.490883
 ENTER SEASONAL MA PARAMETER THETAS(1).
 0.533679

CONVERGENCE HAS BEEN REACHED IN MAX(10,10) ITERATIONS.

SELECTED OUTPUT FOLLOWS:

PARAMETER ESTIMATE	STANDARD ERROR
PHI(1) = 0.146667	0.224464D 00
THETA(1) = 0.509816	0.195934D 00
THETAS(1) = 0.573388	0.784896D-01

MOVING AVERAGE CONSTANT: THETA0 = 0.000248

CHI-SQUARE STATISTIC FOR RESIDUAL LACK OF FIT = 28.302543

DEGREES OF FREEDOM = 37

PROBABILITY OF EXCEEDING STATISTIC = 0.847102

DO YOU WANT TO PLOT AUTO AND PAUTO OF RESIDUALS?

y

*EXECUTION BEGINS...

AUTOCORRELATIONS

-0.013	0.073	-0.100	-0.088	0.077	0.073	-0.031	-0.033	0.105	-0.054
0.033	-0.018	0.014	0.027	0.070	-0.133	0.051	0.008	-0.092	-0.084
-0.020	-0.023	0.208	0.005	-0.036					

PARTIAL AUTOCORRELATIONS

-0.013	0.073	-0.098	-0.096	0.092	0.081	-0.064	-0.040	0.152	-0.050
-0.027	0.017	0.047	-0.004	0.053	-0.118	0.055	0.031	-0.123	-0.135
0.051	-0.006	0.148	-0.019	0.003					

MEAN=0.253052E-02 VARIANCE = 0.139373E-02

ENTER TITLE FOR PLOTS.

autos and pautos of residuals / (1,1,1)x(0,1,1)12 model of series g

YOUR AUTO AND PAUTO PLOTS HAVE BEEN PRINTED OFFLINE.

PICK UP IN ROOM 1140 UNDER YOUR USER ID NUMBER.

DO YOU WANT TO GO AGAIN?

y

```

ENTER LETTER FOR OPTION YOU WANT.
f
IS YOUR DATA IN FILE FT02F001?
y
EXECUTION BEGINS...

IS YOUR SERIES SEASONAL?
y

ENTER SEASONAL DATA AS REQUESTED:

ENTER NUMBER OF SEASONAL DIFFERENCES.
1
ENTER NUMBER OF SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF SEASONAL MA PARAMETERS.
1
ENTER LENGTH OF SEASON VIA FORMAT I2.
12

ENTER NONSEASONAL DATA AS REQUESTED:

ENTER NUMBER OF NON-SEASONAL DIFFERENCES.
1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
1

NOW INPUT YOUR PARAMETER ESTIMATES:

ENTER MA CONSTANT TERM, THETA0.
.000291
ENTER NON-SEASONAL MA PARAMETER THETA(1).
.377152
ENTER SEASONAL MA PARAMETER THETAS(1).
.572387
ENTER MAXIMUM FORECAST LEAD TIME VIA FORMAT I2.
30
ENTER INDEX FOR PLOT ORIGIN VIA FORMAT I3.
100
ENTER INDEX FOR FORECAST ORIGIN VIA FORMAT I3.
131
ENTER SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS.
0.10

WAS YOUR DATA TRANSFORMED IN THE TRANS PROGRAM?
y

DO YOU WANT BASIC OUTPUT AT THE TERMINAL?
y

THE LAST 10 WVEC VALUES:

0.418999E 03  0.461001E 03  0.472000E 03  0.535001E 03  0.622000E 03
0.606000E 03  0.507999E 03  0.461001E 03  0.390001E 03  0.432002E 03

THE 30 FORECAST VALUES:

0.398539E 03  0.415184E 03  0.395402E 03  0.462697E 03  0.450971E 03
0.469710E 03  0.543456E 03  0.618194E 03  0.626276E 03  0.524292E 03
0.460421E 03  0.405213E 03  0.448016E 03  0.466864E 03  0.444748E 03
0.520593E 03  0.507548E 03  0.528791E 03  0.611991E 03  0.696357E 03
0.705666E 03  0.590927E 03  0.519089E 03  0.456978E 03  0.505397E 03
0.526812E 03  0.502002E 03  0.587782E 03  0.573220E 03  0.597386E 03

```

THE 30 UPPER FORECAST CONFIDENCE LIMITS:

0.399598E 03	0.416254E 03	0.396481E 03	0.463785E 03	0.452068E 03
0.470813E 03	0.544567E 03	0.619311E 03	0.627399E 03	0.525422E 03
0.461557E 03	0.406354E 03	0.449173E 03	0.468030E 03	0.445925E 03
0.521779E 03	0.508743E 03	0.529995E 03	0.613203E 03	0.697577E 03
0.706894E 03	0.592163E 03	0.520332E 03	0.458229E 03	0.506664E 03
0.528090E 03	0.503292E 03	0.589083E 03	0.574532E 03	0.598708E 03

ALPHA FOR THE CONFIDENCE LIMITS IS: 0.100

*
DO YOU WANT TO GO AGAIN?
Y

ENTER LETTER FOR OPTION YOU WANT.
f

IS YOUR DATA IN FILE FT02F001?
Y
EXECUTION BEGINS...

IS YOUR SERIES SEASONAL?
Y

ENTER SEASONAL DATA AS REQUESTED:

ENTER NUMBER OF SEASONAL DIFFERENCES.
1
ENTER NUMBER OF SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF SEASONAL MA PARAMETERS.
1
ENTER LENGTH OF SEASON VIA FORMAT I2.
12

ENTER NONSEASONAL DATA AS REQUESTED:

ENTER NUMBER OF NON-SEASONAL DIFFERENCES.
1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.
1
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
1

NOW INPUT YOUR PARAMETER ESTIMATES:

ENTER MA CONSTANT TERM, THETA0.
.000248
ENTER NON-SEASONAL AR PARAMETER PHI(1).
0.146667
ENTER NON-SEASONAL MA PARAMETER THETA(1).
0.509816
ENTER SEASONAL MA PARAMETER THETAS(1).
0.573388
ENTER MAXIMUM FORECAST LEAD TIME VIA FORMAT I2.
30
ENTER INDEX FOR PLOT ORIGIN VIA FORMAT I3.
100
ENTER INDEX FOR FORECAST ORIGIN VIA FORMAT I3.
131
ENTER SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS.
0.10

WAS YOUR DATA TRANSFORMED IN THE TRANS PROGRAM?
Y

DO YOU WANT BASIC OUTPUT AT THE TERMINAL?
Y

THE LAST 10 WVEC VALUES:

0.418999E 03	0.461001E 03	0.472000E 03	0.535001E 03	0.622000E 03
0.606000E 03	0.507999E 03	0.461001E 03	0.390001E 03	0.432002E 03

THE 30 FORECAST VALUES:

0.398650E 03	0.414875E 03	0.395125E 03	0.462366E 03	0.450651E 03
0.469348E 03	0.543045E 03	0.617700E 03	0.625753E 03	0.523897E 03
0.460071E 03	0.404899E 03	0.447739E 03	0.466360E 03	0.444325E 03
0.520096E 03	0.507066E 03	0.528257E 03	0.611381E 03	0.695632E 03
0.704905E 03	0.590336E 03	0.518566E 03	0.456512E 03	0.504960E 03
0.526113E 03	0.501400E 03	0.587074E 03	0.572532E 03	0.596632E 03

THE 30 UPPER FORECAST CONFIDENCE LIMITS:

0.399709E 03	0.415945E 03	0.396204E 03	0.463452E 03	0.451744E 03
0.470448E 03	0.544151E 03	0.618812E 03	0.626870E 03	0.525020E 03
0.461199E 03	0.406032E 03	0.448887E 03	0.467517E 03	0.445491E 03
0.521270E 03	0.508249E 03	0.529448E 03	0.612579E 03	0.696837E 03
0.706117E 03	0.591556E 03	0.519792E 03	0.457745E 03	0.506207E 03
0.527371E 03	0.502668E 03	0.588353E 03	0.573821E 03	0.597930E 03

ALPHA FOR THE CONFIDENCE LIMITS IS: 0.100

DO YOU WANT TO GO AGAIN?

Y

ENTER LETTER FOR OPTION YOU WANT.

F

IS YOUR DATA IN FILE FT02F001?

Y

EXECUTION BEGINS...

IS YOUR SERIES SEASONAL?

Y

ENTER SEASONAL DATA AS REQUESTED:

ENTER NUMBER OF SEASONAL DIFFERENCES.

1

ENTER NUMBER OF SEASONAL AR PARAMETERS.

0

ENTER NUMBER OF SEASONAL MA PARAMETERS.

1

ENTER LENGTH OF SEASON VIA FORMAT I2.

12

ENTER NONSEASONAL DATA AS REQUESTED:

ENTER NUMBER OF NON-SEASONAL DIFFERENCES.

1

ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.

0

ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.

1

NOW INPUT YOUR PARAMETER ESTIMATES:

ENTER MA CONSTANT TERM, THETA0.

.000291

ENTER NON-SEASONAL MA PARAMETER THETA(1).

.377152

ENTER SEASONAL MA PARAMETER THETAS(1).

.572387

ENTER MAXIMUM FORECAST LEAD TIME VIA FORMAT I2.

40

```

ENTER INDEX FOR PLOT ORIGIN VIA FORMAT I3.
075
ENTER INDEX FOR FORECAST ORIGIN VIA FORMAT I3.
120
ENTER SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS.
0.10

WAS YOUR DATA TRANSFORMED IN THE TRANS PROGRAM?
y

DO YOU WANT BASIC OUTPUT AT THE TERMINAL?
n
*
DO YOU WANT TO GO AGAIN?
y

ENTER LETTER FOR OPTION YOU WANT.
d
IS YOUR DATA IN FILE FT02F001?
y
EXECUTION BEGINS...
IS YOUR TIME SERIES SEASONAL?
y
ENTER ORDER OF SEASONAL DIFFERENCING.
1
ENTER LENGTH OF SEASONAL PERIOD VIA I2.
12
ENTER NUMBER OF NONSEASONAL DIFFERENCES.
1
DO YOU WANT TO PLOT AUTO AND PAUTO OF TRANSFORMED DATA?
n
DO YOU WANT TO GO AGAIN?
y

ENTER LETTER FOR OPTION YOU WANT.
c
ENTER DESIRED CP/CMS COMMANDS, ONE PER LINE.
WHEN FINISHED TYPE: &GOTO -QUES
alter file ft02f001 pl insert ft02f001 pl
alter file ft03f001 pl file ft02f001 pl
&goto -ques
DO YOU WANT TO GO AGAIN?
y

ENTER LETTER FOR OPTION YOU WANT.
x
IS YOUR DATA IN FILE FT02F001?
y
EXECUTION BEGINS...

IS YOUR SERIES SEASONAL?
y
ENTER LENGTH OF SEASON VIA FORMAT I2.
12
ENTER NUMBER OF SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF SEASONAL MA PARAMETERS.
1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
1

NOW INPUT YOUR INITIAL PARAMETER ESTIMATES, AS REQUESTED.

ENTER NON-SEASONAL MA PARAMETER THETA(1).
.377152
ENTER SEASONAL MA PARAMETER THETAS(1).
.572387

```

FOR THE GIVEN INPUT PARAMETERS AND MODEL
THE RESIDUAL SUM OF SQUARES IS: 0.18192935D 00

DO YOU WANT TO TEST DIFFERENT PARAMETER VALUES?
n

DO YOU WANT TO TEST A DIFFERENT MODEL?
y

IS YOUR SERIES SEASONAL?
y
ENTER LENGTH OF SEASON VIA FORMAT I2.
12
ENTER NUMBER OF SEASONAL AR PARAMETERS.
0
ENTER NUMBER OF SEASONAL MA PARAMETERS.
1
ENTER NUMBER OF NON-SEASONAL AR PARAMETERS.
1
ENTER NUMBER OF NON-SEASONAL MA PARAMETERS.
1

NOW INPUT YOUR INITIAL PARAMETER ESTIMATES, AS REQUESTED.

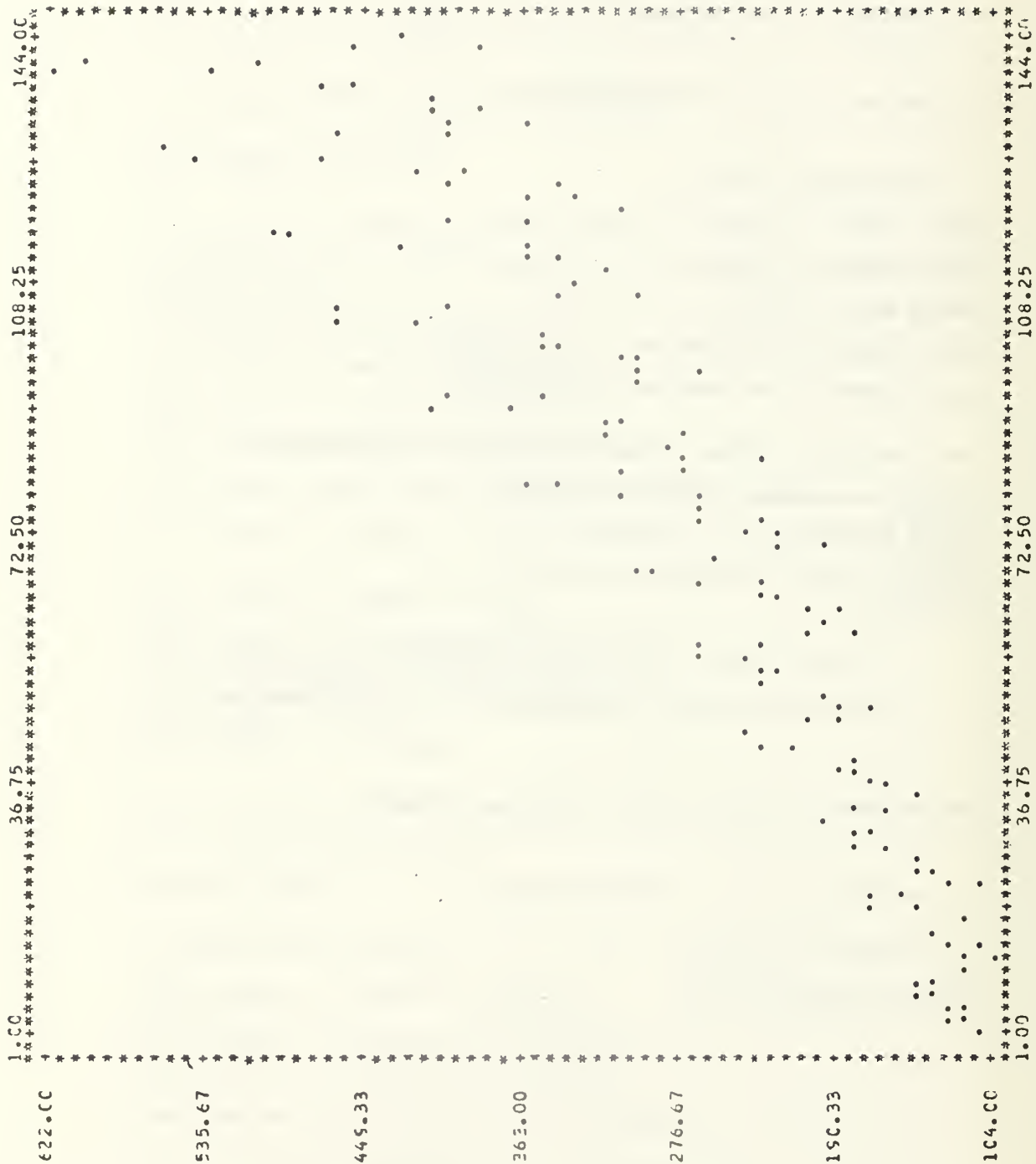
ENTER NON-SEASONAL AR PARAMETER PHI(1).
.146667
ENTER NON-SEASONAL MA PARAMETER THETA(1).
.509816
ENTER SEASONAL MA PARAMETER THETAS(1).
.573388
*

FOR THE GIVEN INPUT PARAMETERS AND MODEL
THE RESIDUAL SUM OF SQUARES IS: 0.18164378D 00

DO YOU WANT TO TEST DIFFERENT PARAMETER VALUES?
n

DO YOU WANT TO TEST A DIFFERENT MODEL?
n

DO YOU WANT TO GO AGAIN?
n
CONTROL RETURNED TO CMS
R;



X-SCALE: "X"= 0.179E 01 UNITS
Y-SCALE: "Y"= 0.863E 01 UNITS
SERIES: G DATA / UNDIFFERENCED

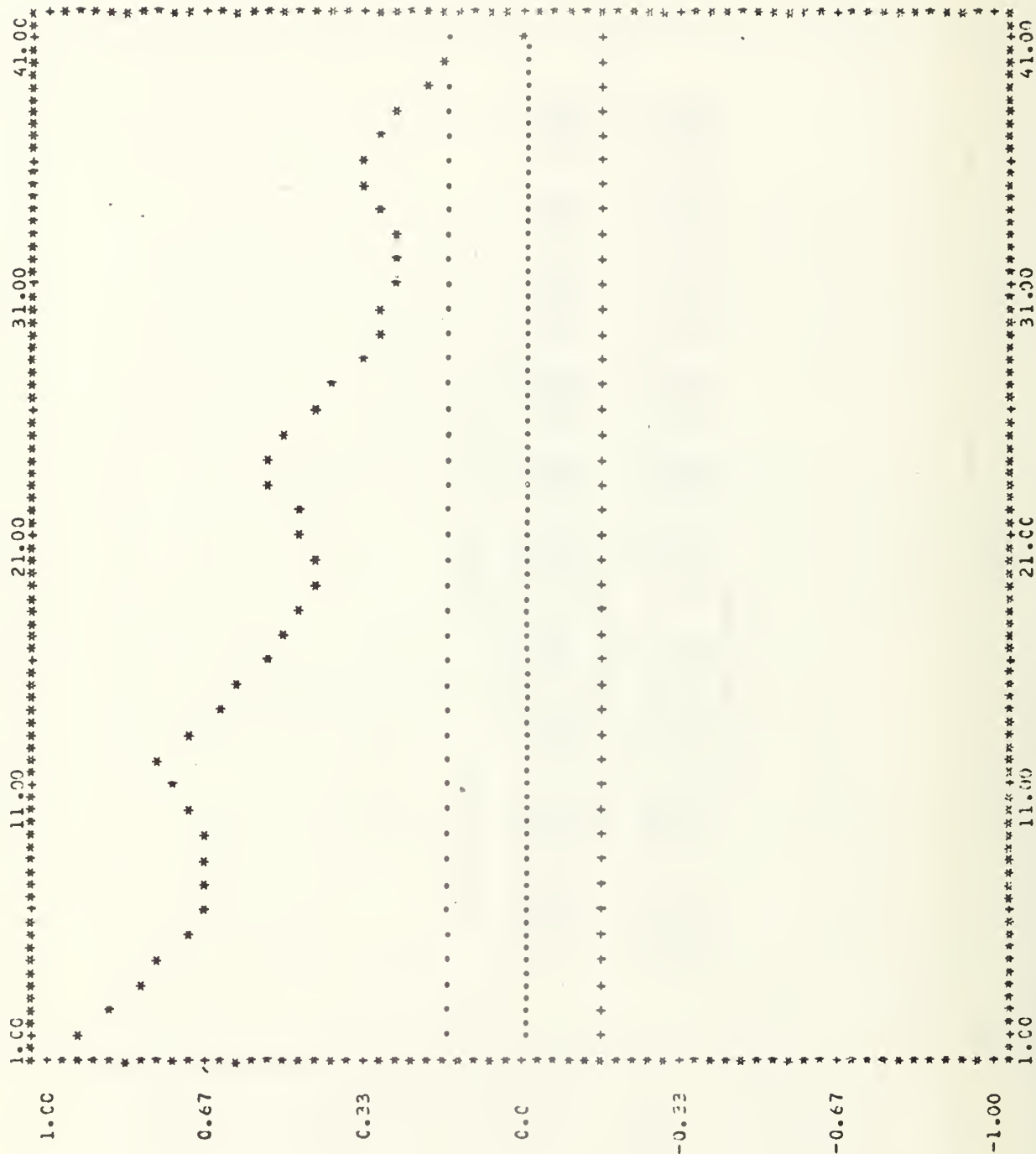
AUTOCORRELATIONS

0.948	C.876	C.807	0.752	0.714	0.682	0.663	0.956	0.671	0.702
0.743	0.760	0.712	0.642	0.586	C.538	0.500	0.465	0.430	0.442
0.457	C.482	0.517	0.522	0.454	C.438	0.388	0.348	0.315	0.288
0.271	0.284	0.277	0.255	0.326	0.337	0.303	0.254	0.211	0.172

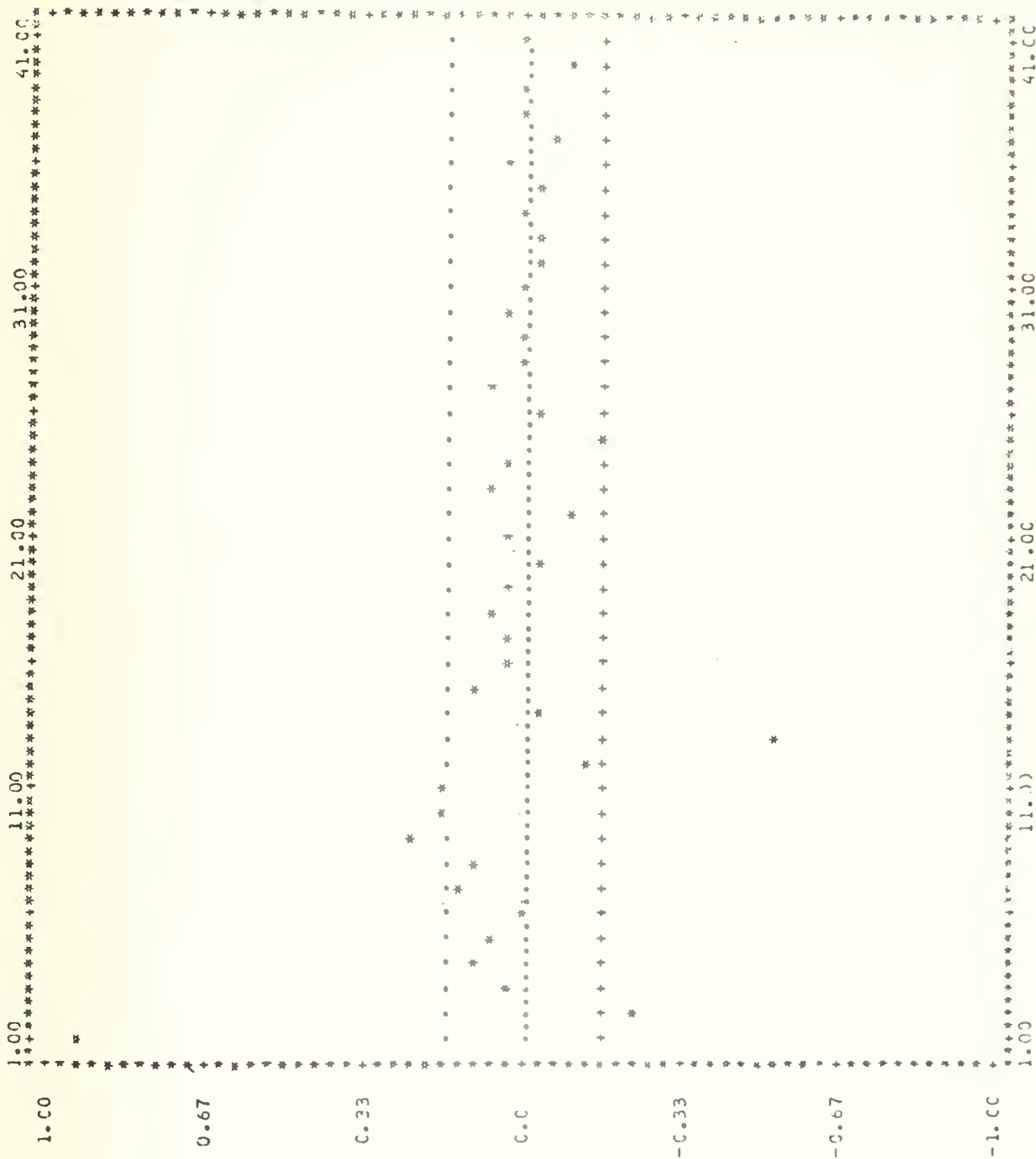
PARTIAL AUTOCORRELATIONS

0.948	-C.225	0.038	0.094	0.074	0.008	0.126	0.090	0.232	0.166
0.171	-0.125	-0.540	-0.227	0.091	0.025	0.032	0.073	0.048	-0.046
0.046	-C.100	0.052	0.048	-0.163	-0.036	0.066	0.006	C.008	-0.019
-0.010	-C.018	-0.025	-0.015	-0.048	0.046	-0.067	-0.002	C.016	-0.088

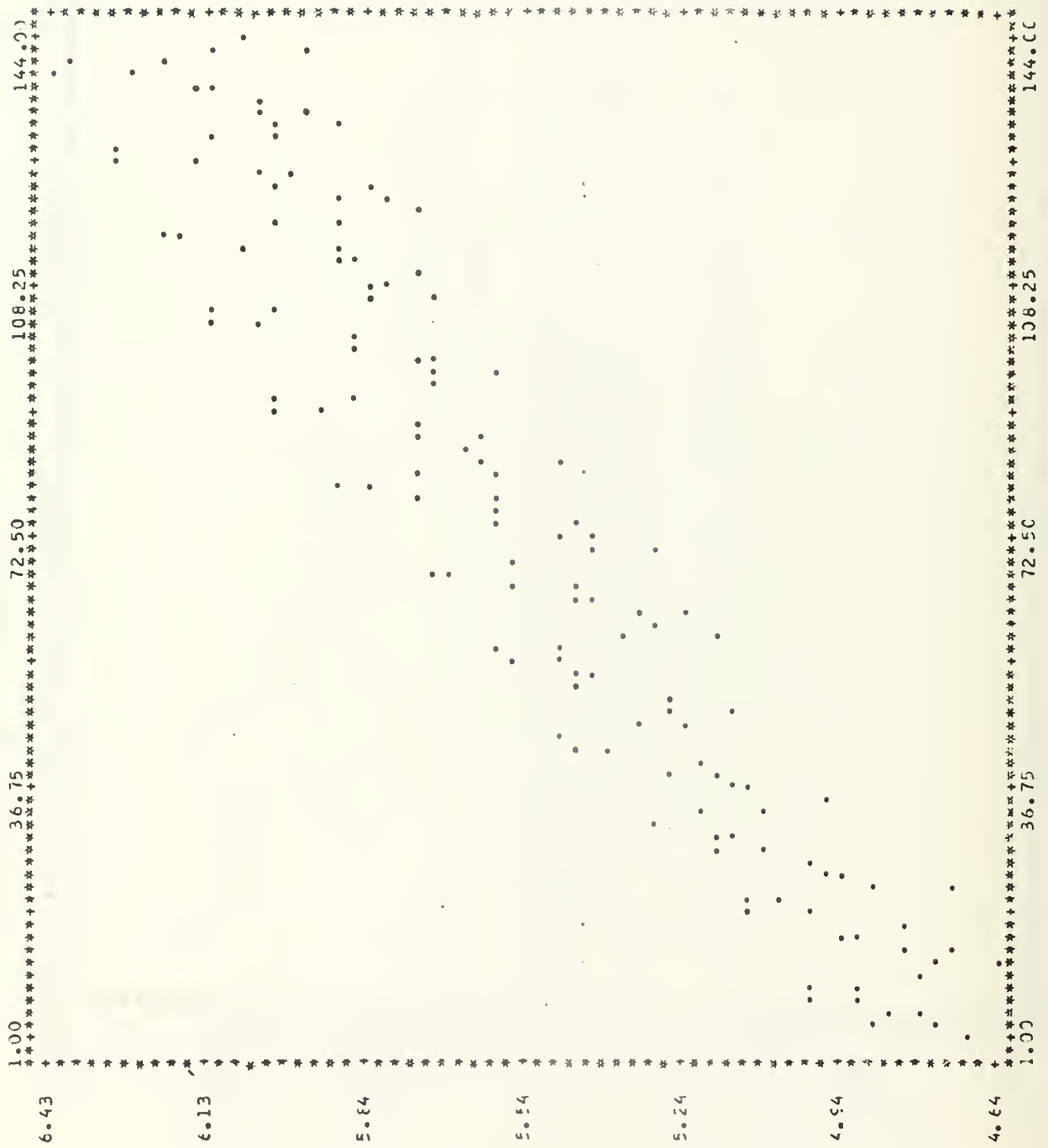
MEAN= 280.299 VARIANCE = 14292.0



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS.
 AUTOS AND PLOTS OF SERIES G DATA / JNCIFFERENCED



X-SCALE: "X" = 0.1793 C1 UNITS
Y-SCALE: "Y" = 0.258E-01 UNITS
PLOT OF SERIES G DATA / NATURAL LOG TRANSFORM

AUTOCORRELATIONS

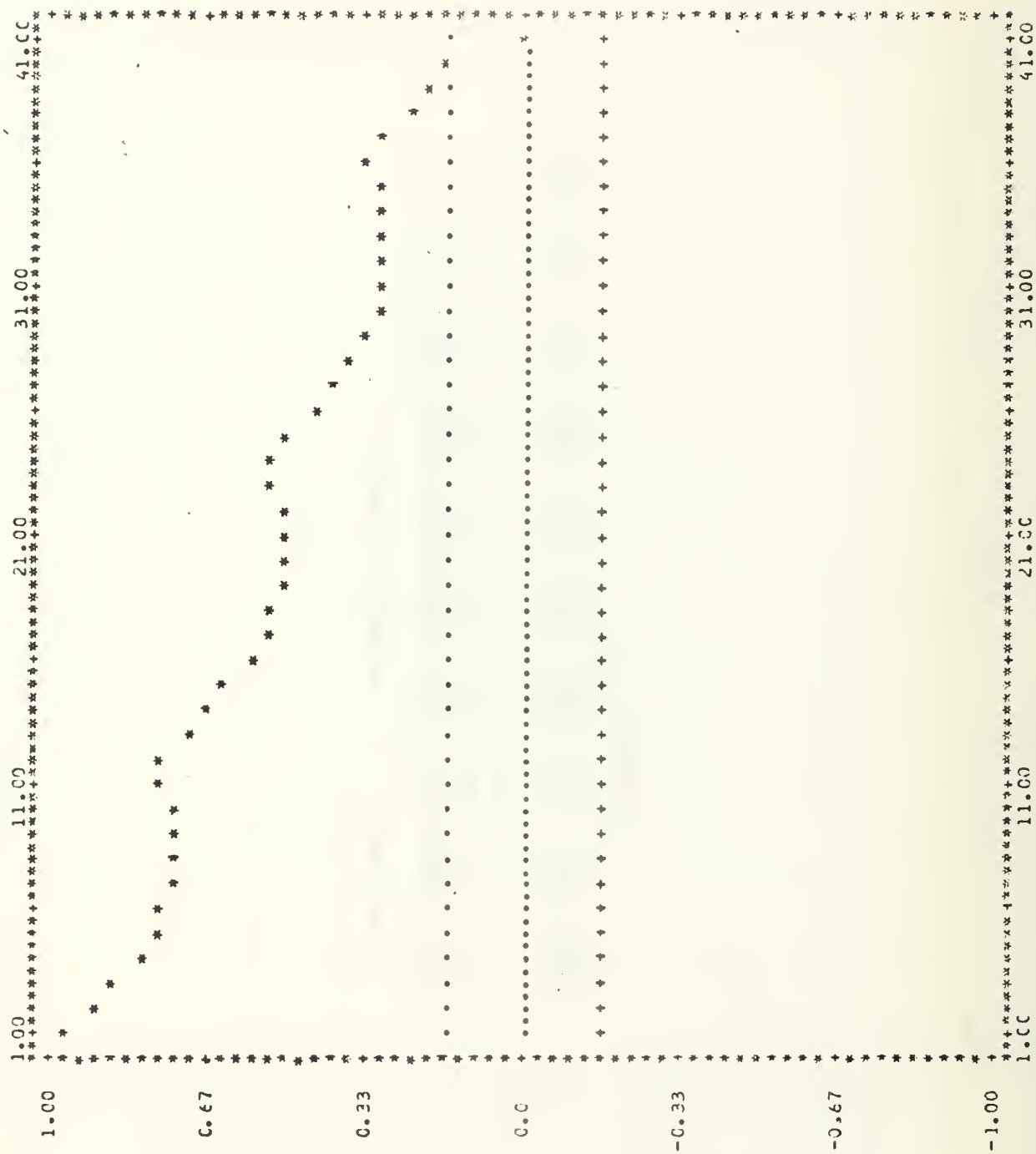
C.954	0.899	0.851	C.808	0.779	C.756	0.738	0.727	0.734	0.744
0.758	0.762	0.717	C.663	0.618	0.576	C.544	0.519	0.531	0.490
0.498	0.506	0.517	C.520	0.484	0.437	0.400	0.364	0.337	0.315
0.257	0.289	C.255	0.305	0.315	0.319	0.286	0.245	0.211	0.175

PARTIAL AUTOCORRELATIONS

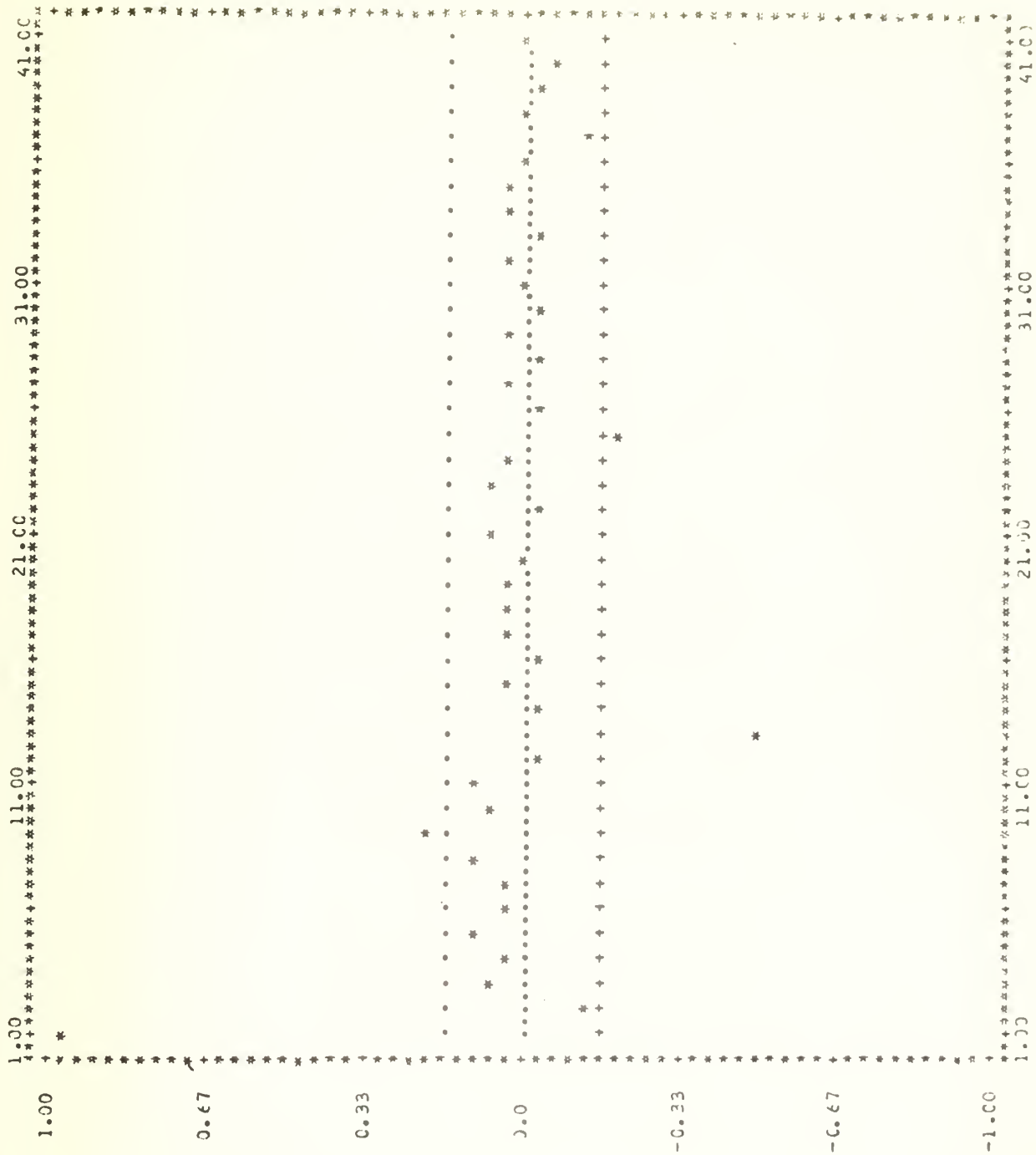
C.954	-C.118	0.054	0.724	0.116	C.044	0.038	0.100	0.204	0.064
0.106	-0.042	-C.485	-C.034	0.042	-0.044	0.028	0.037	0.042	0.014
0.073	-C.033	-0.061	0.031	-0.154	-C.035	0.036	-0.035	C.044	-0.045
-0.003	C.034	-C.020	0.028	0.029	-0.004	-0.132	-0.003	-0.025	-0.055

MEAN= 5.54218

VARIANCE = 0.193531



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS:
AUTOS AND PLOTS OF SERIES G DATA / LOG TRANSFORMED

FILE: FILE	FT03F001 P1	NAVAL POSTGRADUATE SCHCL
0.391703E-01	0.355535E-03	0.661602E-01
0.399103E-01	0.608357E-01	-0.194797E-01
0.751500E-01	0.141555E-01	-0.445404E-01
0.144910E-01	0.234400E-01	-0.516033E-01
0.101459E-01	0.644400E-01	-0.175239E-01
0.194132E-01	0.513000E-01	-0.501085E-01
0.111190E-01	0.292258E-01	-0.363333E-01
0.541100E-01	0.102400E-01	-0.261400E-01
0.304000E-01	0.505338E-01	-0.601100E-01
0.427700E-01	0.849390E-01	-0.404000E-01
0.101000E-01	0.175551E-01	0.335107E-01
0.350500E-01	0.138657E-01	-0.245600E-01
0.121400E-01	0.805500E-01	-0.201597E-01
0.186600E-01	0.235500E-01	-0.215458E-01
0.323500E-01	0.107239E-01	-0.215458E-01
0.238514E-01	0.131440E-01	-0.216100E-01
0.171060E-01	0.240600E-01	-0.222159E-01
0.177250E-01	0.132895E-01	0.339559E-01
0.144550E-01	0.166421E-01	0.415500E-01
0.347200E-01	0.102338E-01	-0.628155E-01
0.130833E-01	0.102338E-01	-0.368155E-01
0.137047E-01	0.455530E-01	-0.500813E-01

AUTOCORRELATIONS

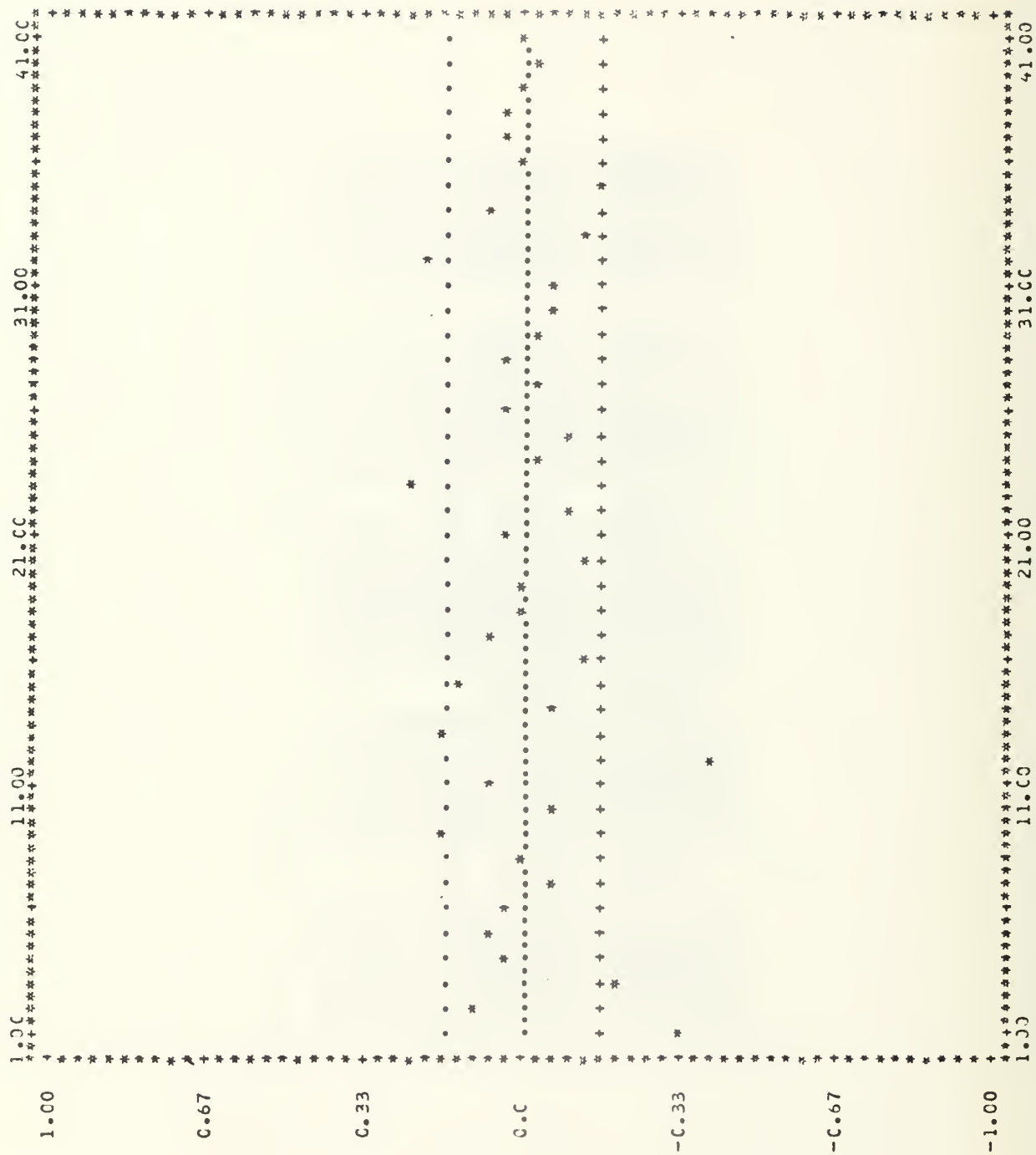
-0.341	0.105	-0.202	0.021	0.156	0.031	-0.056	-0.001	0.176	-0.076
0.064	-0.387	0.152	-0.058	0.150	-0.139	0.071	0.016	-0.011	-0.117
0.039	-0.051	0.223	-0.018	-0.100	0.049	-0.030	0.047	-0.018	-0.051
-0.054	0.196	-0.122	0.078	-0.152	-0.010	0.047	0.031	-0.015	-0.034

PARTIAL AUTOCORRELATIONS

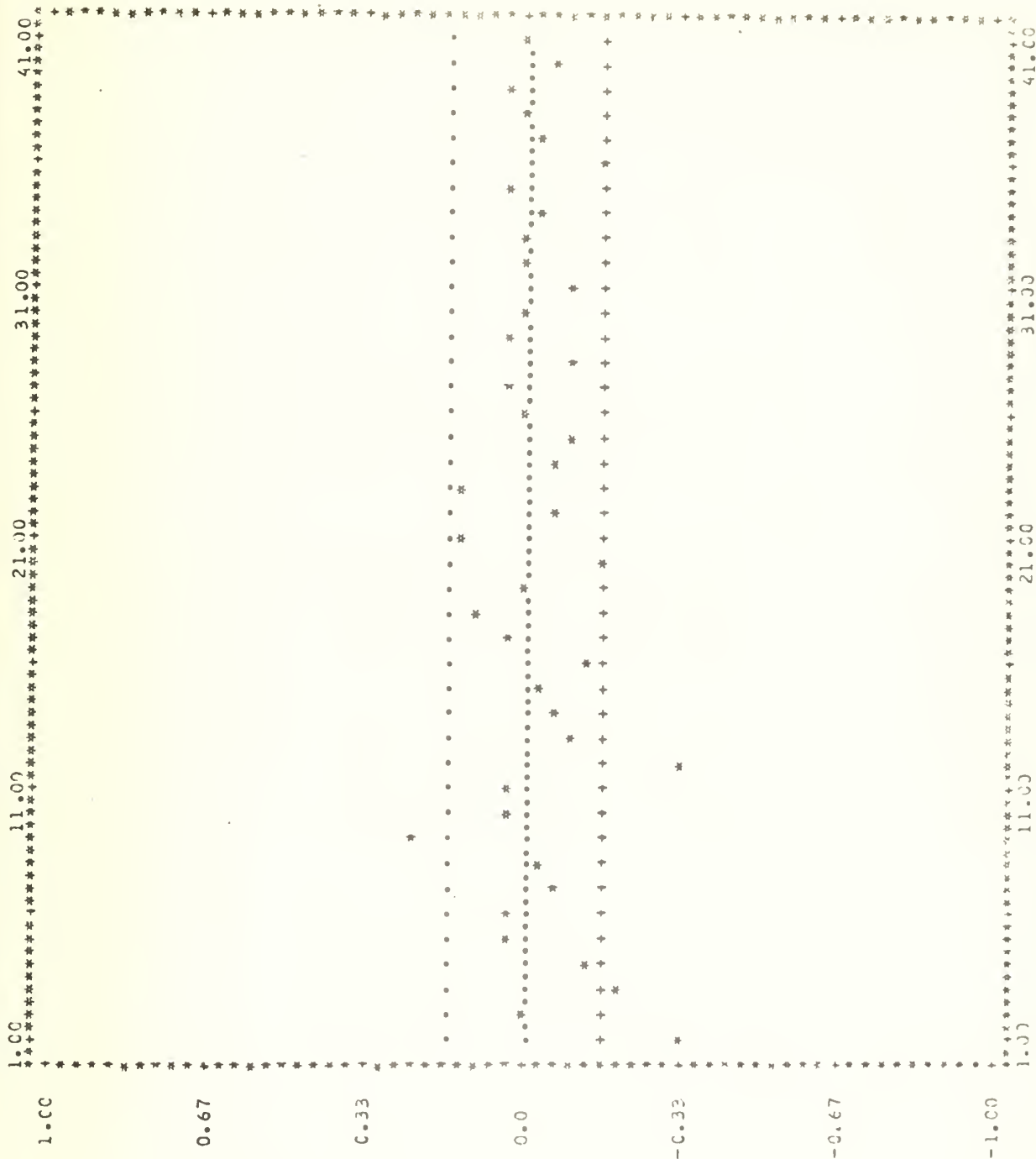
-0.341	-0.013	-0.153	-0.125	0.033	0.035	-0.060	-0.020	0.226	0.043
0.047	-0.339	-0.109	-0.077	-0.022	-0.140	0.026	0.115	-0.013	-0.167
0.132	-0.072	0.143	-0.067	-0.103	-0.010	0.044	-0.050	0.047	-0.005
-0.096	-0.015	0.011	-0.015	0.023	-0.165	-0.034	0.009	0.045	-0.077

MEAN=C.250920E-03

VARIANCE = 0.208604E-02



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS.
 AUTOS AND PLOTS OF SERIES G DATA / 1 MS CLIFF, 1 SEAS CLIFF.

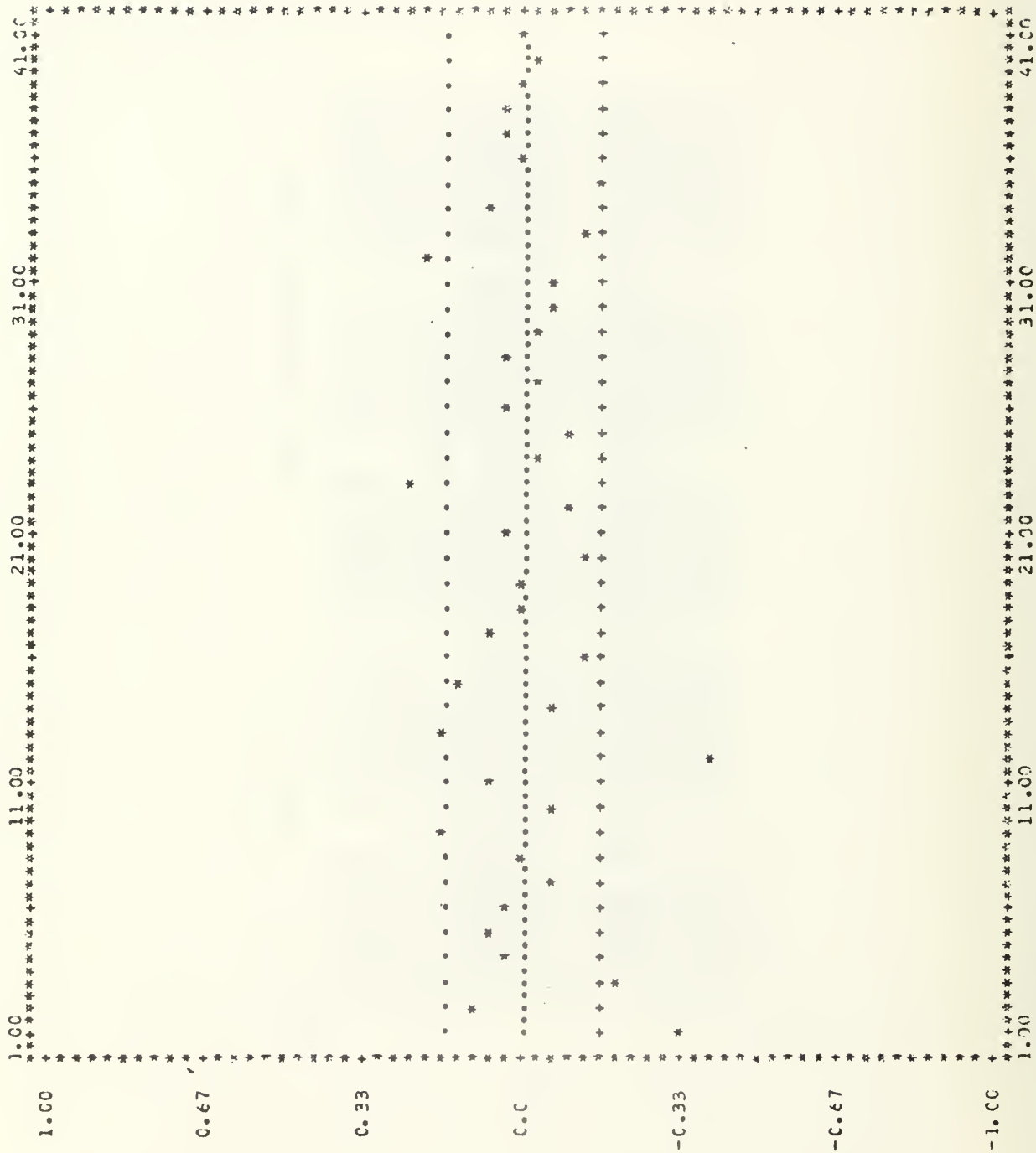
AUTOCORRELATIONS

-0.341	C.105	-C.202	0.021	0.056	-0.056	-0.001	C.176	-0.076
-0.064	-0.387	C.152	-C.058	0.150	-0.071	0.016	-C.011	-0.117
C.035	-C.091	C.223	-C.018	-C.100	-0.033	0.047	-C.018	-0.051
-0.054	C.156	-0.122	0.078	-0.152	0.047	0.031	-C.015	-0.034

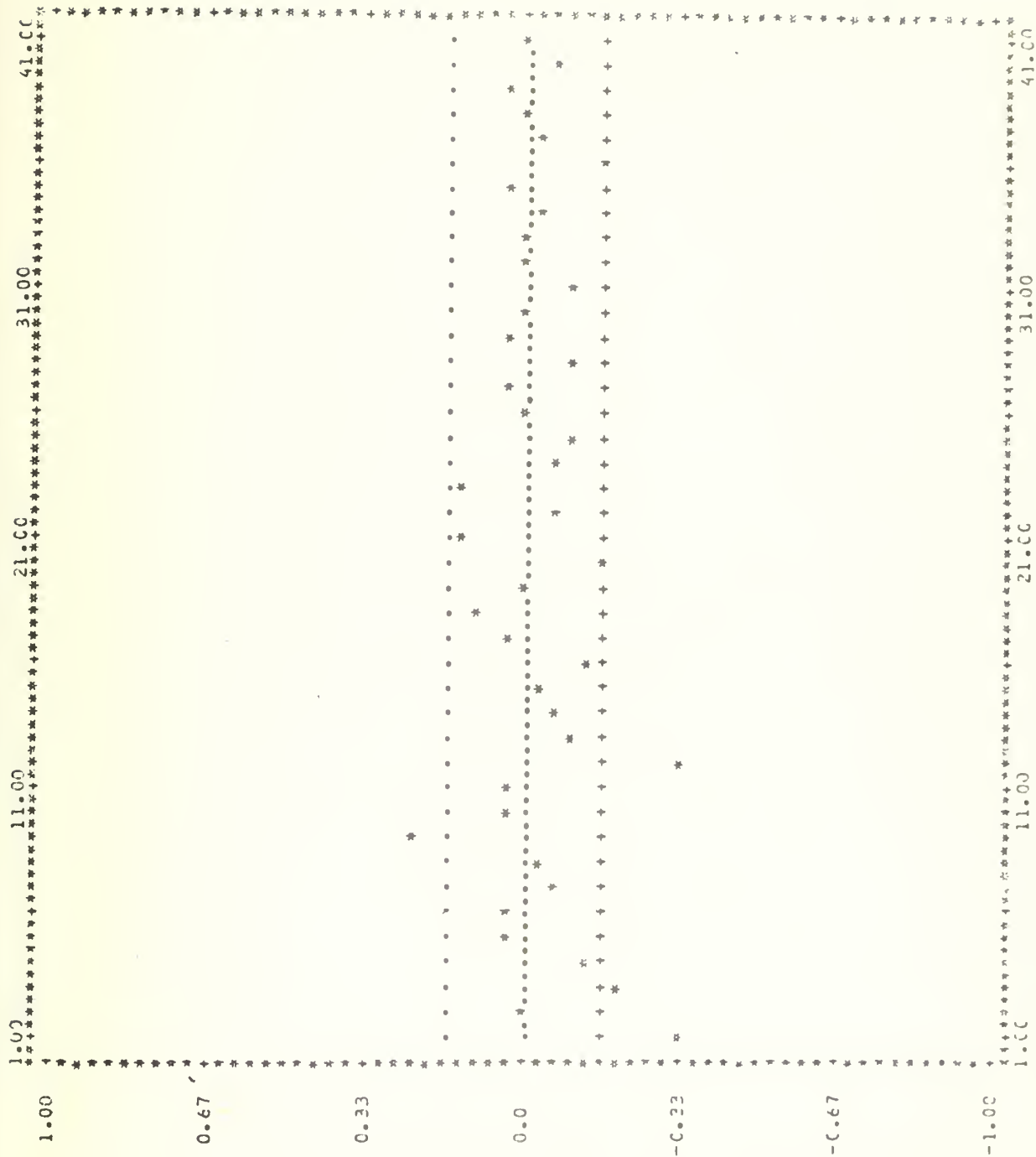
PARTIAL AUTOCORRELATIONS

-0.341	-0.013	-0.193	-0.125	0.033	-0.035	-0.020	0.126	0.043
0.047	-0.339	-0.105	-0.077	-0.022	-0.140	0.026	-0.013	-0.167
C.132	-0.072	0.143	-0.067	-0.103	-C.010	0.044	0.047	-0.003
-0.096	-0.015	0.011	-C.015	0.023	-C.165	-0.034	0.045	-0.077

MEAN=C.290920E-03 VARIANCE = C.208604E-02



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS.
 ACTUS AND PACTUS FOR SERIES G DATA / 1 NS DIFF, 1 SEAS DIFF.

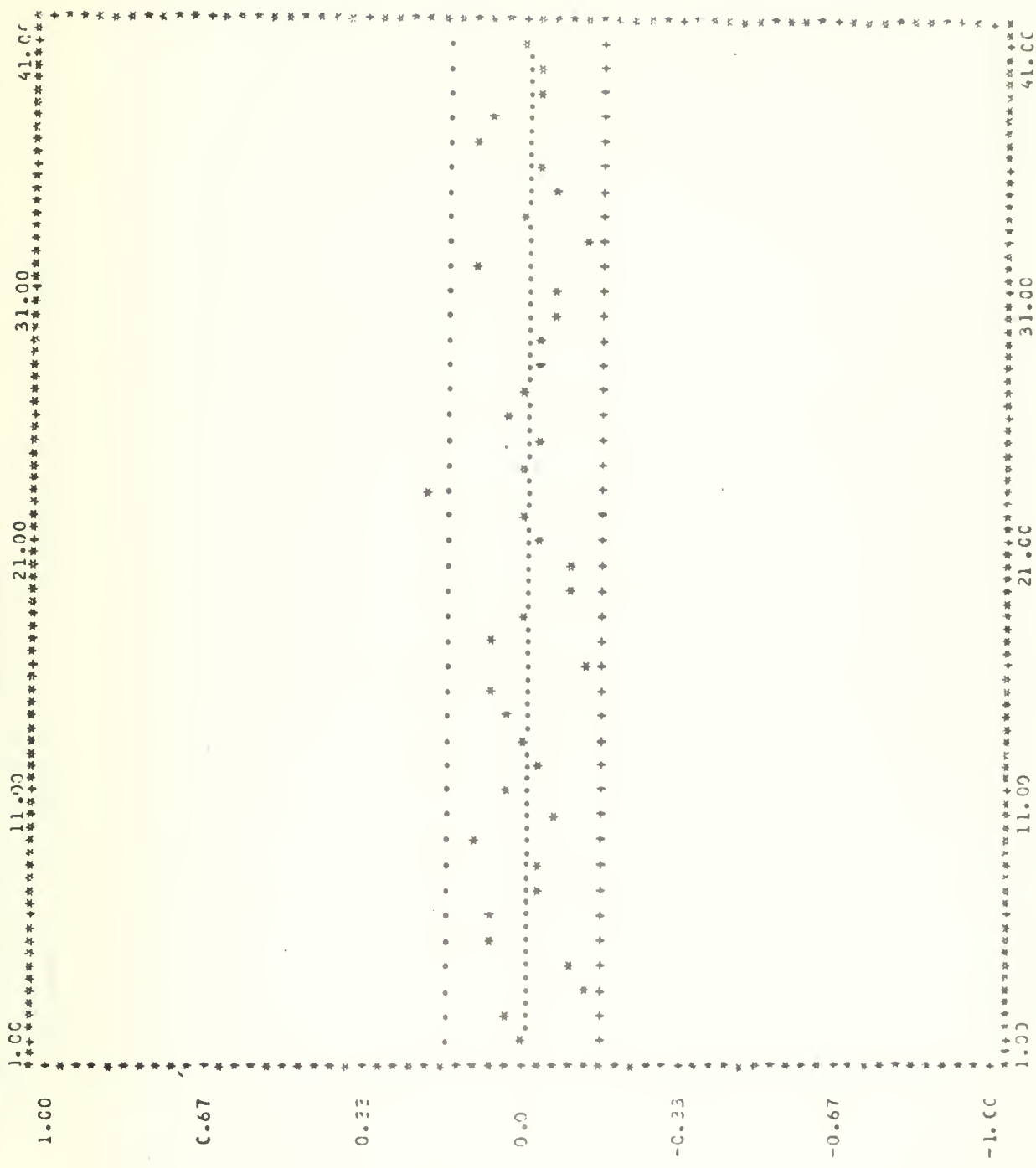
AUTOCORRELATIONS

C.005	-C.129	-C.105	0.078	C.077	-0.036	-0.036	-0.033	C.103	-0.050
-0.026	C.020	C.035	0.067	-0.130	-0.047	-0.052	-0.015	-0.093	-0.050
-0.027	-C.114	C.013	-0.042	-0.036	-0.036	-0.011	-0.032	-0.127	-0.074
-0.063		-0.142	-0.081			C.089	0.062	-0.024	-0.035

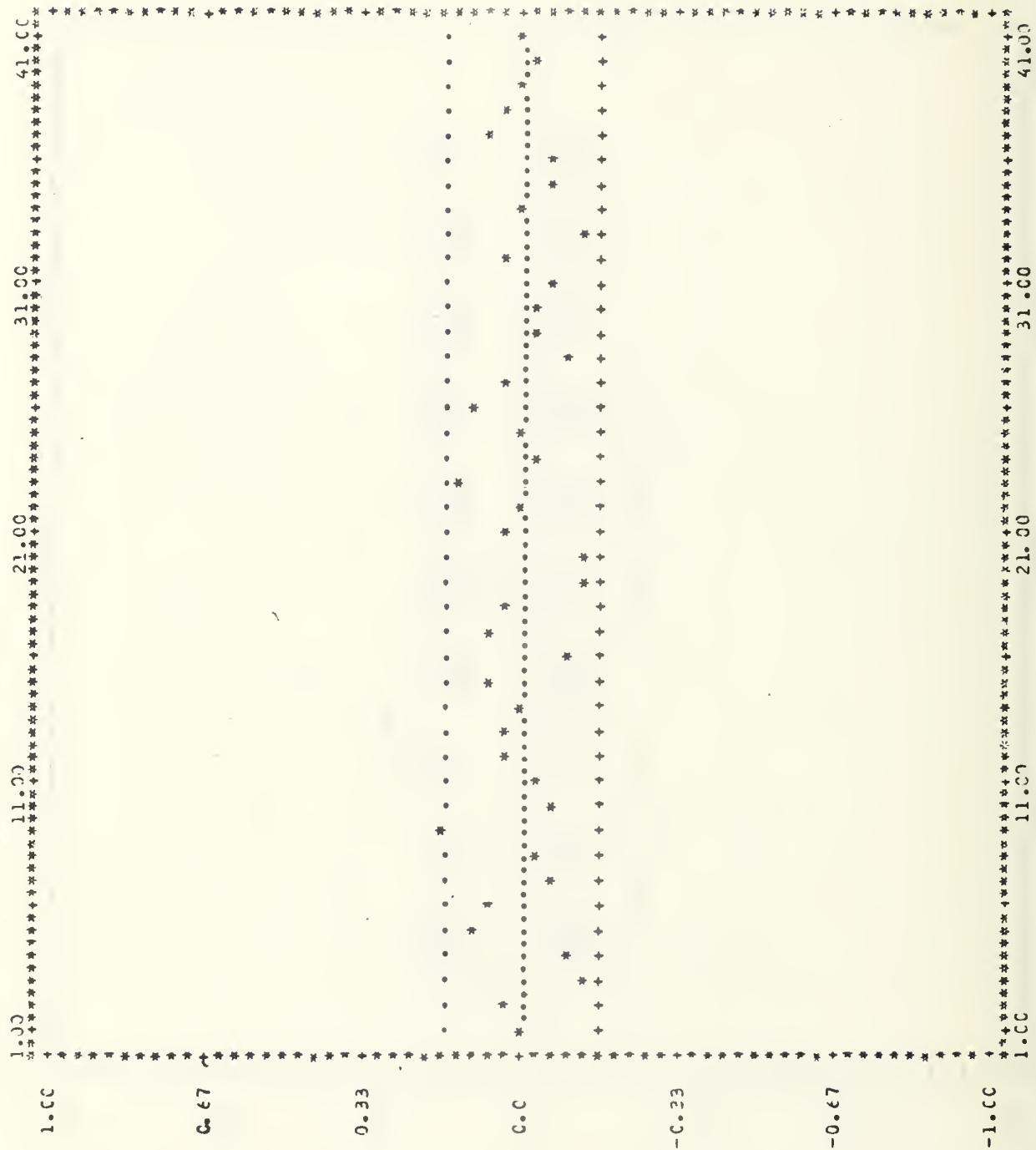
PARTIAL AUTOCORRELATIONS

C.005	0.026	-0.130	-0.104	0.068	-0.072	-0.029	0.151	-0.057
-0.022	0.018	0.046	0.053	-0.114	0.069	0.029	-0.129	-0.134
-0.048	-0.014	-0.143	0.013	0.068	0.024	-0.091	-0.023	-0.033
	0.025	-0.125	-0.074	-0.074	0.082	0.041	-0.010	-0.042

MEAN=C.236005E-02 VARIANCE = 0.139713E-02



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS.
 AUTOS AND PLOTS OF RESIDUALS / (0,1,1)X(0,1,1)12 MODEL OF S

AUTOCORRELATIONS

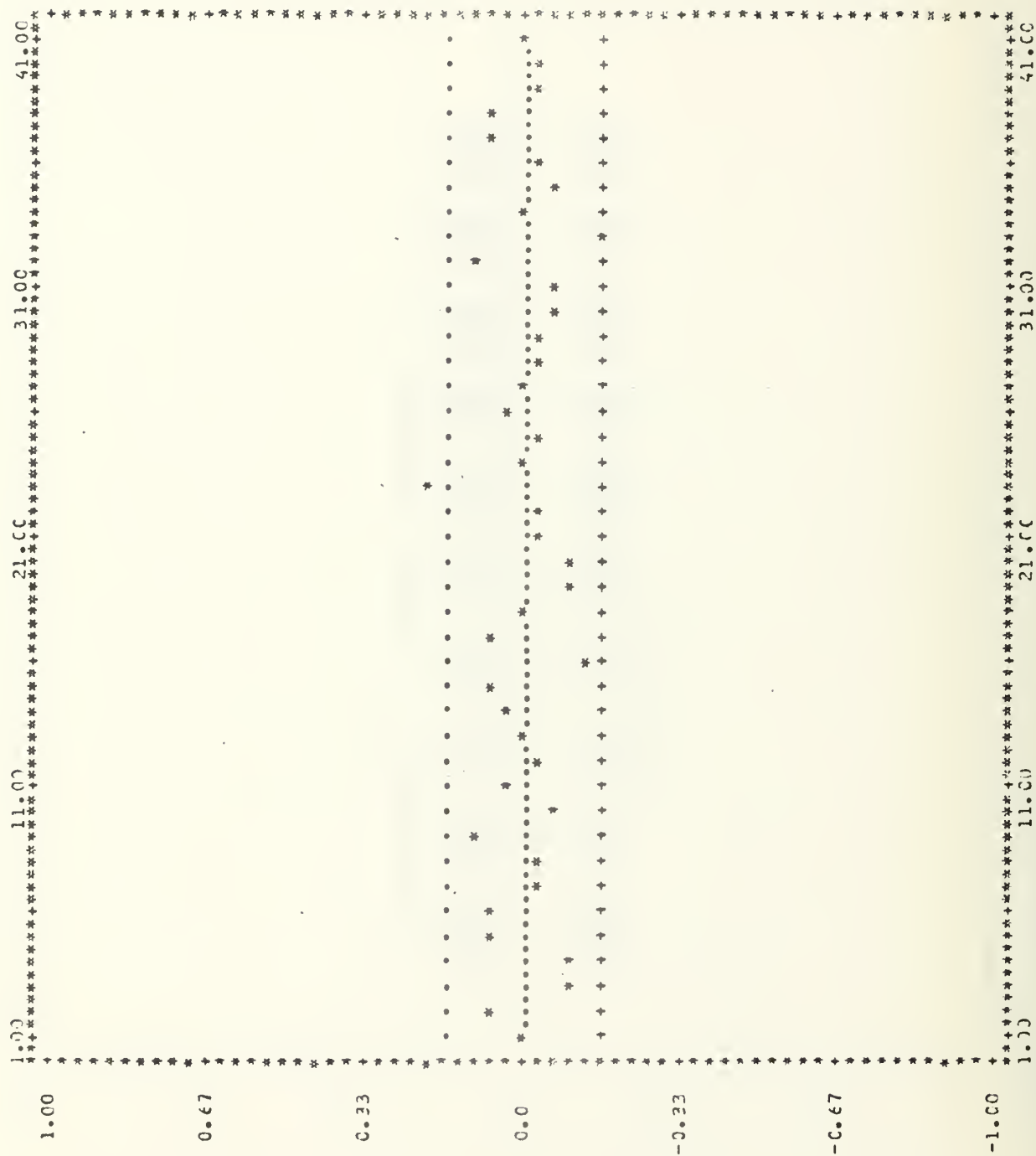
-0.013	-0.075	-0.100	-0.088	0.077	0.073	-0.231	-0.033	0.125	-0.054
-0.033	-0.018	0.014	0.027	0.070	-0.133	0.051	0.008	-0.092	-0.084
-0.020	-0.023	0.208	0.005	-0.036	-0.048	-0.014	-0.034	-0.027	-0.073
-0.014	0.111	-0.154	0.197	-0.182	-0.037	0.080	0.052	-0.027	-0.032

PARTIAL AUTOCORRELATIONS

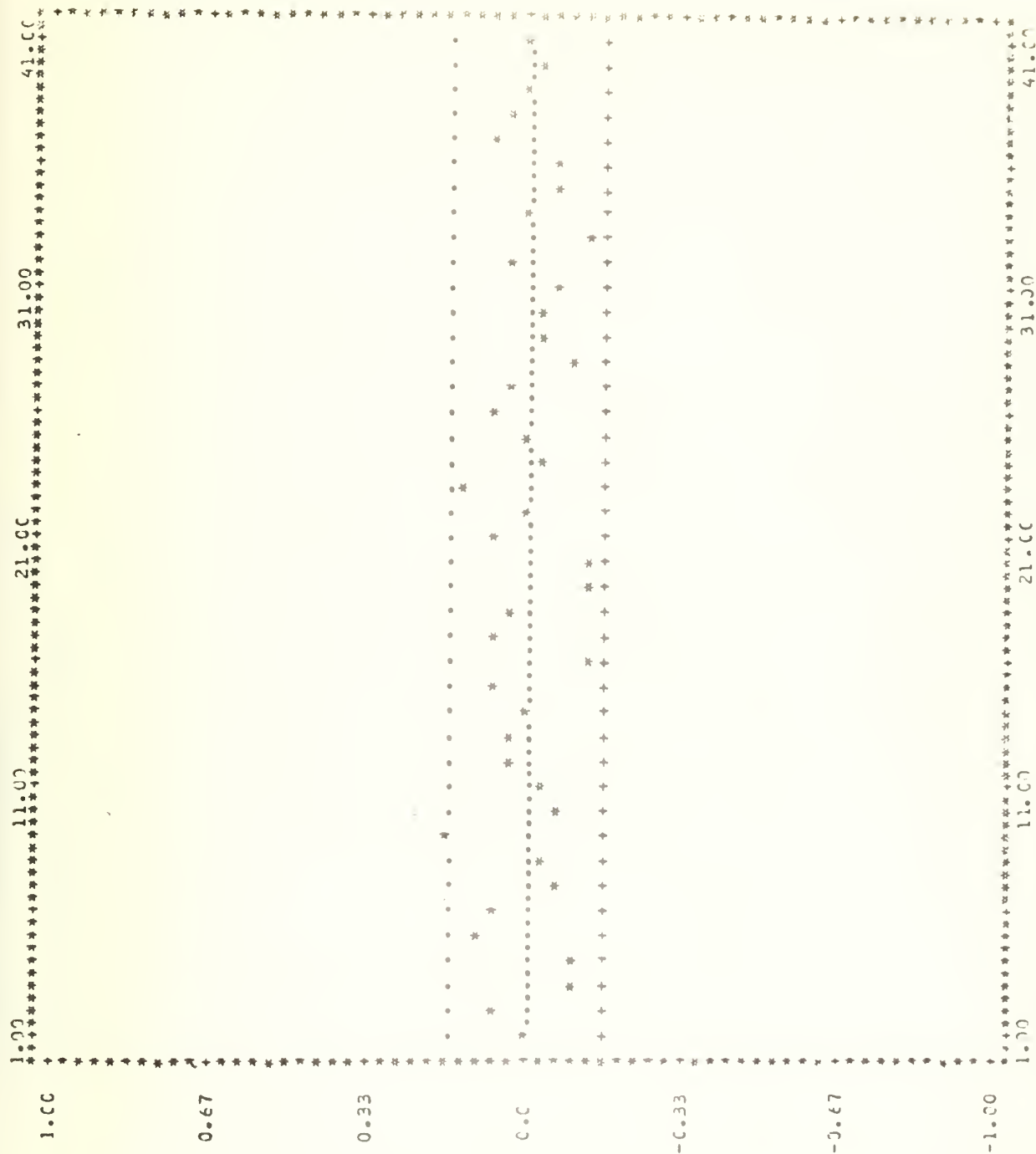
-0.013	0.073	-0.098	-0.056	0.092	0.081	-0.264	-0.040	0.152	-0.050
-0.027	0.017	0.047	-0.004	0.053	-0.118	0.055	0.031	-0.123	-0.135
-0.051	-0.006	0.148	-0.015	0.003	0.079	0.062	-0.099	-0.031	-0.032
-0.063	0.027	-0.130	0.008	-0.062	-0.072	0.082	0.044	-0.012	-0.044

MEAN=0.253052E-02

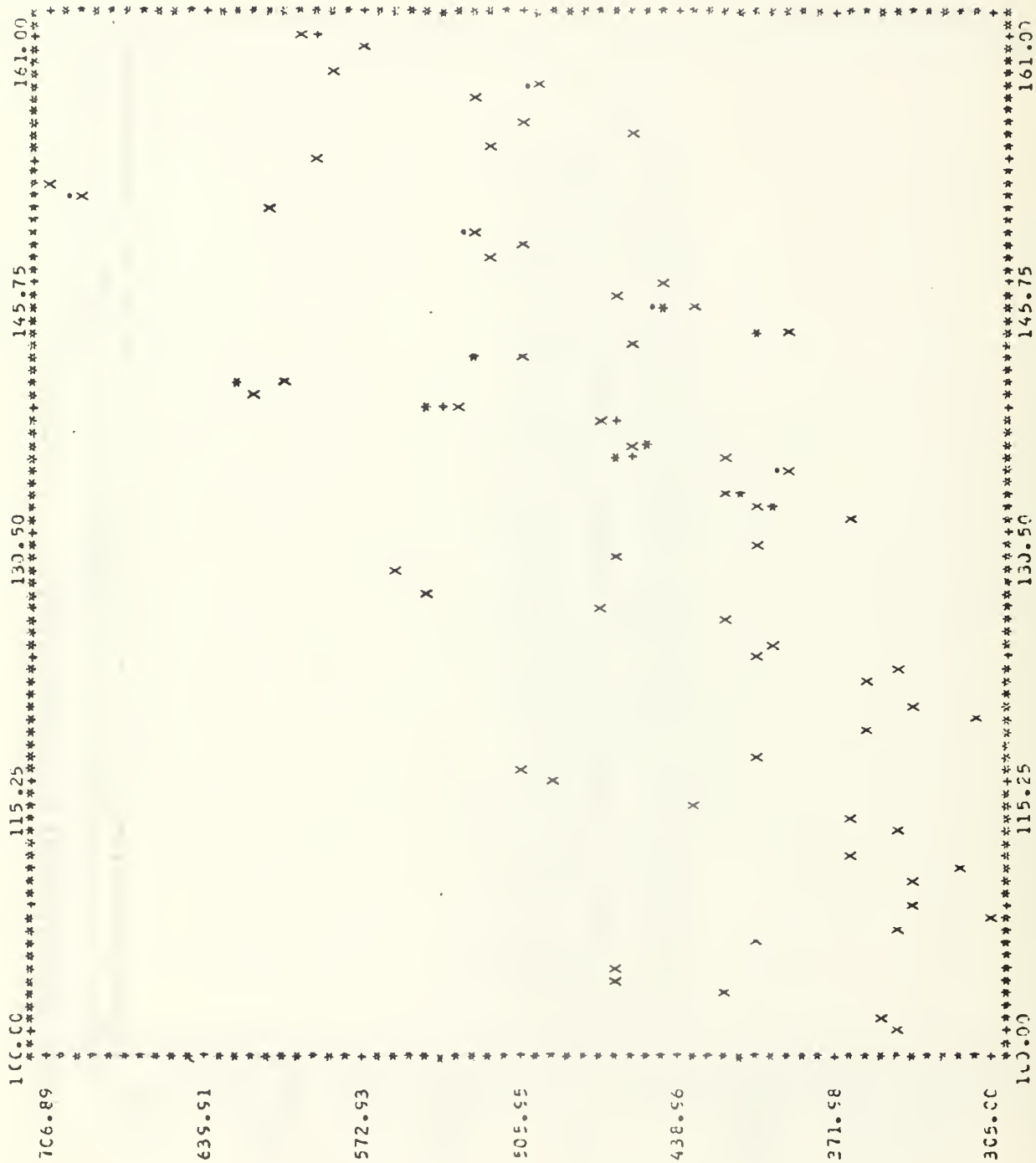
VARIANCE = 0.139373E-02



AUTOCORRELATIONS WITH 2 SIGMA BANDS.



PARTIAL AUTOCORRELATIONS WITH 2 SIGMA BANDS.
 AUTOS AND PLOTS OF RESIDUALS / (1,1,1)(0,1,1)12 MODEL OF S



FORECAST VALUES:

398.55063	415.134012	395.401611	462.695333	450.571191
466.709717	543.456055	618.193604	626.273835	524.272491
460.421387	403.212048	448.016357	466.963225	444.748291
520.553018	507.548096	528.791260	611.991455	590.356334
705.666016	590.526758	519.088867	456.578271	505.557177
526.811523	502.002441	587.781982	573.219571	597.335742

UPPER CONFIDENCE LIMITS:

399.598339	416.254255	396.481445	463.784512	452.661627
470.812477	544.566955	619.311035	627.399412	525.422503
461.557129	406.354248	449.173196	468.030273	445.924805
521.779053	508.742320	529.994373	613.203269	697.577148
706.854043	592.192598	520.332275	458.229248	506.663818
528.090083	503.252236	589.083008	574.531738	598.708252

LOWER CONFIDENCE LIMITS:

397.275492	414.113525	394.321533	461.607910	449.874512
468.605713	542.344571	617.075928	625.151611	523.162354
459.285400	404.771045	446.859275	465.656933	443.171533
519.406982	506.353027	527.587402	610.779257	695.136475
704.437744	589.690674	517.845215	455.727651	504.103771
525.532955	500.712462	586.480713	571.907555	596.662588

SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS: 90.000

FORECAST ORIGIN:131

MAXIMUM FORECAST LEAD TIME: 30



ORIGINAL AND FORECASTED TIME SERIES.

* UPPER CONFIDENCE INTERVAL
+ LOWER CONFIDENCE INTERVAL
* FORECASTS

FORECAST VALUES:

398.649658	414.874512	395.124756	462.365723	450.650879
469.348389	543.045166	617.699707	623.352086	523.826729
460.000557	407.839170	477.739258	460.359883	444.224707
520.009570	507.006162	528.257324	611.381348	695.632380
704.000027	590.236426	518.566466	450.512451	504.959717
526.112545	501.235658	587.073975	572.532227	596.632324

UPPER CONFIDENCE LIMITS:

399.708984	415.645068	396.203857	463.452393	451.742305
470.448486	544.151367	618.811768	626.870361	525.019775
461.158730	406.032471	448.886963	467.517304	445.450967
521.270264	508.148115	523.472510	617.579103	696.837158
724.111432	592.255604	515.702236	457.745117	506.207275
527.571094	502.608457	588.352539	573.820557	597.530176

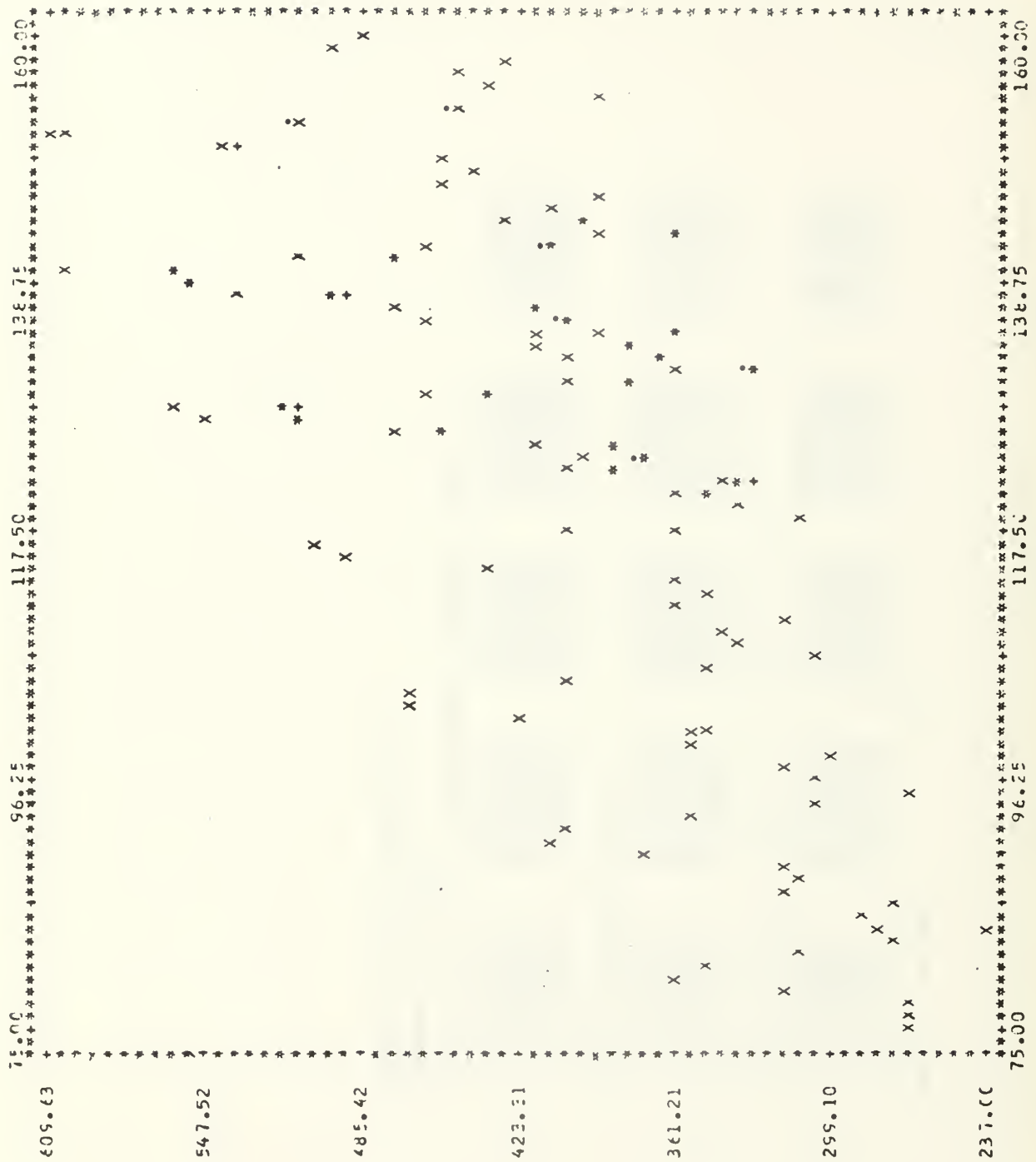
LOWER CONFIDENCE LIMITS:

397.590088	413.802711	394.045410	461.270809	449.557129
426.248047	541.038121	616.587402	624.634766	522.773438
438.542130	403.785625	446.591309	465.202148	443.158203
518.520898	505.883545	527.066855	617.183350	694.426758
703.652871	589.116543	517.340322	455.279541	503.111514
524.853760	500.130615	585.795166	571.243652	595.334229

SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS: 90.000

FORECAST ORIGIN: 131

MAXIMUM FORECAST LEAD TIME: 30



FORECAST VALUES:

349.764404	320.545166	385.547598	375.771725	385.523228
456.098389	510.597114	514.459229	433.913086	380.518125
322.142090	368.207164	378.895010	361.421554	418.427777
407.413036	418.541148	474.791260	554.183833	518.425475
471.135010	413.244971	360.843750	471.142334	411.445475
392.527813	455.202666	443.238379	455.503682	538.445640
603.475342	608.273926	513.338379	450.502682	393.435640
436.366943	445.281582	428.822021	496.626513	483.554346

UPPER CONFIDENCE LIMITS:

50.325195	334.617188	397.030229	376.862545	387.015949
457.205078	511.811523	515.579834	433.020283	381.711677
333.641454	365.252027	380.045058	362.401554	419.155433
478.604248	419.542189	496.002488	555.401554	515.155433
472.365025	414.587649	362.054238	401.471554	413.155433
394.291250	456.301023	444.572598	456.824701	535.175003
604.818113	609.627157	514.701904	451.876463	354.780047
437.760742	450.652848	430.248251	458.067383	483.708477

LOWER CONFIDENCE LIMITS:

248.703355	332.472900	384.865722	374.680664	384.824463
454.551455	509.533740	513.338379	432.755645	375.444336
331.003441	367.062256	377.723377	352.440222	417.101116
406.211660	417.341064	492.581787	552.565572	557.201416
466.460146	412.172051	350.593018	398.384353	410.590120
551.703234	453.774102	441.953125	454.182373	537.311779
602.132234	606.520410	511.974609	449.128662	392.012270
434.572530	447.805873	427.395508	455.186275	482.500000

SIGNIFICANCE LEVEL FOR CONFIDENCE INTERVALS: 90.000

FORECAST CRICIN:120

MAXIMUM FORECAST LEAD TIME: 40

APPENDIX C

LISTING OF DATA SETS RESIDING IN THE TIME SERIES EDITOR

This appendix contains listings of the data sets now residing in the Time Series Editor. The data consists of Box and Jenkins' [Ref. 4] time series C [Ref. 4, p. 528] and time series G [Ref. 4, p. 531], as well as a data set containing monthly Monterey, California, rainfall data from January 1931 through December 1976 .

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89</											

FILE: FAIN

[illegible]

LIST OF REFERENCES

1. Anderson, O.D., Time Series Analysis and Forecasting; The Box-Jenkins Approach, Butterworths, Boston, 1976.
2. Anderson, T.W., The Statistical Analysis of Time Series, John Wiley & Sons, New York, 1970.
3. Ansley, C.F., Spivey, W.A., and Wroblek, W.J., A Class of Transformations for Box-Jenkins Seasonal Models, Applied Statistics, 16, No. 2, p. 173, 1977.
4. Box, G.E.P. and Jenkins, G.M., Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.
5. Chatfield, C. and Prothero, D.L., Box-Jenkins Seasonal Forecasting; Problems in a Case-study, Journal of the Royal Statistical Society, A, 136, Part 3, p. 295, 1973.
6. Control-Program 67/Cambridge Monitor System (CP-67/CMS) User's Guide, IBM Corporation, White Plains, New York, 1972.
7. Cox, D.R. and Lewis, P.A.W., The Statistical Analysis of Series of Events, John Wiley & Sons, New York, 1966.
8. Draper, N.R. and Smith, H., Applied Regression Analysis, John Wiley & Sons, New York, 1966.
9. The IMSL Library, Volume 1, International Mathematical and Statistical Libraries, Houston, Texas, 1975.
10. Jenkins, G.M. and Watts, D.G., Spectral Analysis and its Applications, Holden-Day, San Francisco, 1968.
11. Lee, W.L., An Application of CP/CMS to Time Series Analysis, Master's Thesis, Naval Postgraduate School, Monterey, 1977.
12. Mabert, V.A., An Introduction to Short Term Forecasting Using the Box-Jenkins Methodology, American Institute of Industrial Engineers, Inc., Norcross, Georgia, 1975.
13. Marquardt, D.W., An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial Applied Mathematics, Volume 11, No. 2, June, 1963.

14. Nelson, C.R., Applied Time Series Analysis for Managerial Forecasting, Holden-Day, San Francisco, 1973.
15. Pindyck, R.S. and Rubinfeld, D.L., Econometric Models and Economic Forecasts, McGraw-Hill, New York, 1976.
16. Wheelwright, S.C. and Mauridakis, S., Forecasting Methods for Management, John Wiley & Sons, New York, 1977.
17. Wheelwright, S.C. and Mauridakis, S., Interactive Forecasting, Univariate and Multivariate Methods, Holden-Day, San Francisco, 1978.

INITIAL DISTRIBUTION LIST

No. Copies

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
Library, Code 55 Naval Postgraduate School Monterey, California 93940	1
Naval Postgraduate School Monterey, California 93940	
Attn:	
Professor F. Russell Richards, Code 55Rh	50
Professor P. A. W. Lewis, 55Lw Code 55Lw	1
Professor P. W. Zehna, 55Ze Code 55Ze	1
Professor D. R. Barr, Code 55Bn	1
Professor A. R. Washburn, Code 55Ws	1
Professor D. P. Gaver, Code 55Gv	1
Professor P. A. Jacobs, Code 55Jc	1
Professor P. R. Milch, Code 55Mh	1
Professor R. H. Shudde, Code 55Su	1
Professor G. T. Howard, Code 55Hk	1
Professor T. Jayachandran, Code 53Ja	1
Professor R. S. Elster, Code 54Ea	1
Professor P. M. Carrick, Code 54Ca	1
LCDR S. R. Woodall, USN 711 South Skyline Plaza 3705 South George Mason Drive Falls Church, Va. 22041	5
Dr. Kneale T. Marshall Pers Or, Bureau of Naval Personnel Arlington, Annex Washington, D.C. 20370	1
R. J. Stampfel, Code 55 Naval Postgraduate School Monterey, California 93940	1

U185864

DUDLEY KNOX LIBRARY - RESEARCH REPORTS



5 6853 01057689 5

~~U185864~~